FRAGMENT-BASED DRUG DESIGN FOR PROSTATE CANCER

HALİSE BÜŞRA ÇAĞIRICI

KOC UNIVERSITY
AUGUST 2015

Fragment-Based Drug Design for Prostate Cancer

by

Halise Büşra Çağırıcı

A Thesis Submitted to the
Graduate School of Science and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Biomedical Sciences and Engineering

Koç University

August 2015

Koc University

Graduate School of Sciences and Engineering

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

Halise Büşra Çağırıcı

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

Committee Members:

_____

Metin Türkay, Ph. D. (Advisor)

_____

İ.Halil Kavaklı, Ph. D.

_____

Ersin Acar, Ph. D.

Date: _____

ABSTRACT

Drug design process requires a high experimental cost and takes several years for a drug candidate to become commercially available. Computer-aided studies in the drug development have been emerged to reduce cost and time spent in both synthesis and screenings of drug candidates. Fragment-based drug design is one of the approaches in computer-aided studies. In this approach, small fragments binding to the target site are linked together, in silico, to cover the target site and optimize the complementarity. Currently computational approaches have been applied in the development of drugs against many diseases including AIDS, hypertension, and prostate cancer.

Prostate cancer which is the second leading cause of death among men in Turkey is the focus of this thesis. High levels of serum androgen are highly associated with prostate cancer. Common treatment options either cause serious side effects or end in drug resistivity. Yet specific inhibition of an enzyme, CYP17, is an intriguing research subject since it would exert control over androgen biosynthesis without serious side effects.

In this thesis, a fragment-based drug design approach is employed in order to develop novel inhibitors against CYP17 protein in the treatment of prostate cancer. A scaffold for drug candidates is determined based on a lead compound. Fragments suitable for the scaffold are determined and combined in order to increase specificity and the affinity of the lead. Molecular dynamics simulation and docking tools are used to evaluate inhibitory effects of small compounds and GAMS is used to identify novel inhibitors for CYP17 by calculating combinatorial effects of individual fragments.

ÖZET

İlaç tasarlama süreci yüksek bir maliyet gerektirir ve olası bir ilaç molekülünün piyasaya sürülmesi yıllar sürer. İlaç moleküllerinin sentezi ve taranması için gereken maliyet ve zamanın düşürülmesi için ilaç endüstrisinde bilgisayar destekli çalışmalar kullanılmaya başlanmıştır. Parça bazlı ilaç tasarımı da bilgisayar temelli çalışmaların bir yöntemidir. Bu yöntemde, hedef bölgeyi kaplayabilmek ve en kuvvetli bütünleşmeyi sağlayabilmek için hedef bölgeye bağlanan küçük parçalar bilgisayar ortamında birbirine bağlanır. Şu anda hesaplamalı yöntemler AIDS, yüksek tansiyon ve prostat kanserinin de aralarında bulunduğu birçok hastalığa karşı ilaç geliştirilmesinde kullanılmaktadır.

Bu tez çalışmasında Türkiye'deki erkek ölümlerinin en yaygın ikinci nedeni olan prostat kanseri çalışışmıştır. Kanda bulunan yüksek androjen seviyesi prostat kanseriyle oldukça bağlantılıdır. Yaygın olan tedavi seçenekleri ciddi yan etkiler taşımaktadır veya ilaca karşı dirençle karşılaşılmaktadır. Buna rağmen, CYP17 enziminin işlevinin durdurulması androjen biyosentezi mekanizmalarının ciddi yan etkiler olmadan kontrol edilmesini sağlayabileceğinden, bu enzime özel bir inhibisyon oldukça dikkat çeken bir araştırma konusudur.

Bu tez çalışmasında, parça bazlı ilaç dizaynı metodu kullanılarak prostat kanseri tedavisine yönelik olarak CYP17 proteinine özel yeni inhibitörlerin bulunması amaçlanmıştır. İlaç adayları için önceden belirlenen öncü molekülü doğrultusunda bir yapı iskeleti oluşturulmuştur. Öncü molekülden daha spesifik ve daha uyumlu ilaç adaylarının tasarlanması için bu yapı iskeleti doğrultusunda küçük parçalar belirlenmiş ve birleştirilmiştir. Moleküler modelleme simülasyonları ve ilaç tarama programları kullanılarak ilaç adaylarının enzime bağlanma kuvvetleri ölçülmüştür ve GAMS programı ile her bir parçanın toplamdaki etkisi hesaplanarak yeni CYP17 enzimi inhibitör molekülleri belirlenmiştir.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitudes to Dr. Metin Türkay and Dr. İ. Halil Kavaklı for their patience, guidance and support during my graduate study. Also I would like to thank my thesis committee member Dr. Ersin Acar for participating in my thesis committee and his support in my thesis defence. I owe my special thanks to my life partner Emre Çağırıcı for always believing in me and for his encouragement. It would not be possible finish this thesis without his support and patience during this process. Finally I would like to thank all of my big family for being there for me at all the time and for their care in my wellbeing. This thesis was financed by the Scientific and Technological Research Council of Turkey (TUBITAK).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Drugs are chemical substances acting on physiological process to either inhibit or enhance activity of a target protein according to the desired treatment outcome. The first drugs were discovered from roots or leaves of plants by trial and error processes. Currently, drug discovery is a multi-phase process whereby basic knowledge of pathophysiology obtained through academic discoveries is translated into a medical entity that can be used to improve human health [1].

With the advances in science and technology, drug discovery have gain attention from both academia and industry, and have become a scientific challenge where several disciplines including structural biology, chemistry and computational techniques are merged [2]. Chemistry-related disciplines offer vast number of drug candidates to be screened and this large set of drug candidates can be tested experimentally by high-throughput screening (HTS). Yet, applications of HTS technology require a high cost, and specificity or the complementarity of the drug candidates to their target proteins/systems cannot be assessed.

Computational approaches in the drug development have gained importance since these applications can reduce cost and time spent in both synthesis and screenings of drug candidates [3]. Advances in molecular biology and crystallography techniques enable virtual visualization of the protein structures, which is a milestone in drug development research. Protein structures, active sites, and interaction sites can be visualized and

molecular mechanisms of protein-protein or protein-ligand interactions can be studied with these techniques. Drug candidates targeting interaction site or active site of the target protein can be designed and can be screened virtually to eliminate vast number of drug candidates to be tested experimentally using computational drug design techniques. As they help to eliminate the number of drug candidates and to increase specificity of drug candidates to the target protein, computer-based drug design can significantly simplify the process with decreasing the cost and the time. Once a drug candidate, called lead compound, with micro-molar affinity for the target protein is identified after those steps, the lead is modified in order to enhance its affinity to nano-molar scales and to eliminate its undesirable properties such as toxicity or insolubility, in which the process called lead optimization

Recent computer-based drug development approaches divided into at least three categories: virtual screening, inspection, and de novo drug design [4]. In the first category, virtual screening, commercially available chemicals stored in large databases are docked into target protein in silico and scored based on the algorithms to predict binding energy of the ligand. In the second category, inspection, the structure of known drug molecules that bind the target site is modified to become more potent inhibitors/activators and to increase its binding affinity to the target by maximizing the complementary interactions in the target site. In the final category, de novo drug design, small fragments binding to the target site are linked together, in silico, to cover the target site and optimize the complementarity. There are various algorithms/tools for computer-based drug development approaches, which will be discussed in Overview chapter.

Currently computational approaches in drug development have been applied in the treatment many diseases including AIDS, hypertension, and prostate cancer. Prostate cancer is the focus of this thesis. Cancer is the first leading cause of death worldwide [5] where prostate cancer is the second leading cause of death among men in Turkey [6].

Common treatment options include chemotherapy, radiotherapy, surgical procedures and hormonal therapy. However, all of the treatment options either cause serious side effects or end in drug resistivity. Androgens are the major growth factors for prostate cell growth and high levels of serum androgen levels are associated with prostate cancer. Biosynthesis of the potent androgens requires a key enzyme, CYP17. Specific inhibition of this enzyme is an intriguing research subject since it would exert control over androgen biosynthesis without serious side effects.

In this thesis, a computational de novo drug design, fragment-based drug design, approach has been employed in order to develop novel inhibitors against CYP17 protein. A scaffold for drug candidates were determined based on a lead compound discovered by Armutlu et al. [7]. Fragments suitable for the scaffold were determined and combined in order to increase specificity and the affinity of the lead. Molecular dynamics simulation and docking tools are used to evaluate inhibitory effects of small compounds and GAMS is used to identify novel inhibitors for CYP17 by calculating combinatorial effects of individual fragments.

Chapter 2 is devoted to the necessary background information and literature review for prostate cancer and the computer-based drug design research so far. Chapter 3 is dedicated to the computational methods used and the proposed methodology suggested in this study. Chapter 4 provides the results of computational tools and discussions for the proposed methodology. Finally, this thesis is concluded with a summary of the performed study and future work in Chapter 5.

Chapter 2

Overview

2.1. Prostate Cancer

2.1.1 Background Information for Prostate Cancer

According to World Health Organization (WHO), in 2011, cancer is the first leading cause of death worldwide, responsible for nearly 8 million of the total of 55 million deaths from all causes (~14% of all deaths around the world were caused by cancer [5]) as shown in the Figure 2.1. WHO data over a 30-year period also demonstrates that prostate cancer (PC) is the second most frequently diagnosed cancer among men worldwide.

1 out of 7 men are diagnosed with PC in their lifetime [8]. As shown in Figure 2.2, PC is the second leading cause of cancer among men in Turkey, according to the Turkish Ministry of Health [6], which is consistent with the WHO data.

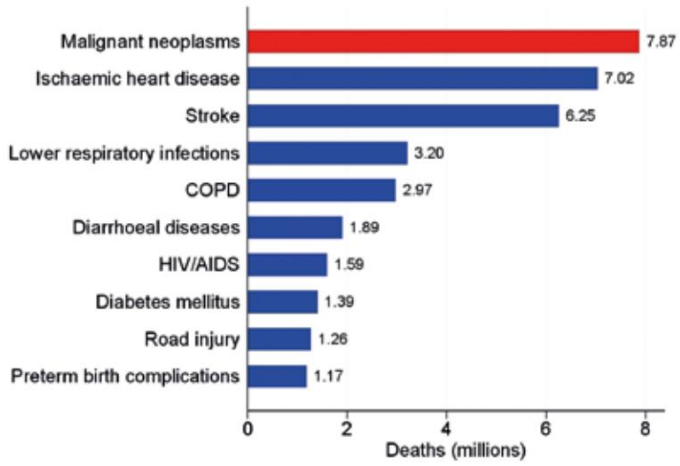


Figure 2.1: the 10 leading causes of death worldwide in 2011 (COPD, chronic obstructive pulmonary disease) [5]

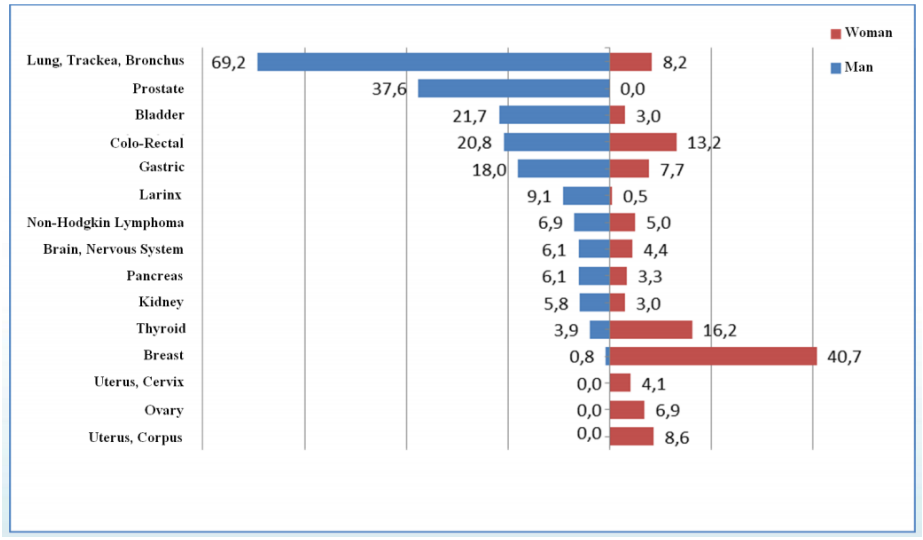


Figure 2.2: Cancer incidence rates in Turkish population by 2005 [6]

Treatment options for PC include radiotherapy, surgery, androgen deprivation therapy (ADT) etc. Radiotherapy and surgery act directly on the affected cells/tissues to

either kill the cancerous cells or remove all or part of the malignant tissue. ADT aims to reduce the levels of androgens by either surgical castration or LHLH agonists/antiandrogens for androgen biosynthesis.

As shown in Figure 2.3, androgens lead to cell proliferation through recognition via androgen receptors (ARs) and AR downstream pathways. Cholesterol or adrenal androgens are converted to testosterone and dihydrotestosterone (DHT) by the enzyme CYP17. Serum testosterone is also converted to DHT, which is a more potent androgen, by the enzyme 5α-reductase [9]. Recognition of DHT or other androgens via AR in malignant and stromal cells initiates transcriptional activation and release of growth factors (GFs). Src and related Src family kinases can interact with GF receptors in order to activate downstream signaling, which results in cell proliferation and survival. Thus, androgen ablation may induce cellular apoptosis of malignant cells.

Androgen deprivation therapy (ADT) has two mechanisms of action either through surgical castration or the use of luteinizing hormone-releasing hormone (LHLH) agonists and antiandrogens. Several studies have focused on ADT in the treatment of PC and there are clinically approved LHLH agonists/antiandrogens acting directly on the testicular production of androgens [8, 10, 11] Although initially effective, majority of the patients will eventually develop castration-resistant prostate cancer (CRPC), which is fatal in the most of patients [8]. Possible mechanisms behind CRPC development include androgen biosynthesis from other sources accelerated to induce AR signaling and drug resistivity developed against antiandrogens [8, 12, 13]. Since resistance against drugs may arise and CRPC may progress after castration, there is a continuing need for the new alternative therapeutics that can slow the progression of prostate cancer.

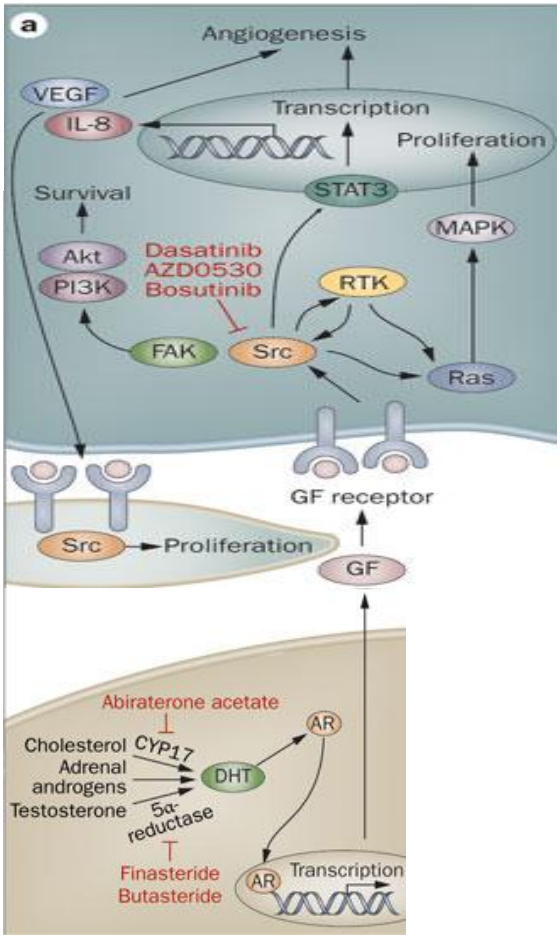


Figure 2.3: Androgen signaling. Adopted from Ref. [9]. Serum testosterone can be converted to the more potent androgen DHT. Once binding of testosterone or DHT to AR, androgens initiate transcriptional activation and release of growth factors (GFs). The downstream signaling induces cell proliferation. Drugs that target specific targets of androgen signaling pathway are shown in red.

Other than targeting AR to block interaction of androgens with the androgen receptors, androgen biosynthesis pathway is of importance in the treatment of PC. Figure 2.4 shows a more detailed representation of human testosterone biosynthesis pathway. Androgen synthesis pathway starts from cholesterol and includes many downstream processes catalyzed by several enzymes till the formation of testosterone and DHT [14].

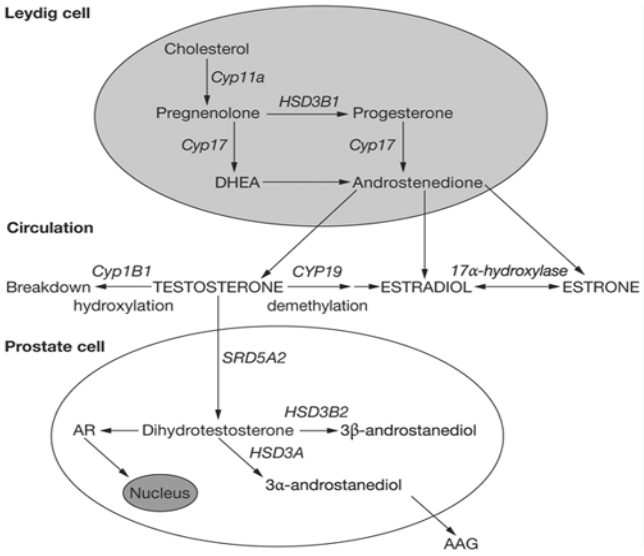


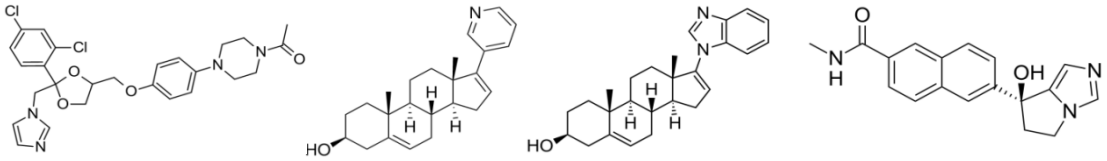
Figure 2.4: Human testosterone biosynthesis pathway [14]. Biosynthesis process starts from cholesterol till DHT formation.

The cytochrome P450 family of enzymes are involved in the maintenance of body homeostasis by metabolizing endogenous hormones and their biochemical intermediates. Cyp17 is a key enzyme in the process of androgen synthesis since it is involved in conversion of progesterone to androstenedione and pregnenolone into dehydroepiandrosterone (DHEA) and which is intermediate key compound for testosterone.

2.1.2 CYP17 Inhibitors in clinical use and clinical trials

In spite of low levels of serum androgens in castrated men, studies show upregulation of androgen receptor and key enzymes in AR signaling [15, 16]. This finding is also supported by upregulation of CYP17 in CRPC, indicating an active androgen biosynthesis pathway in CRPC [17]. Thus therapies that inhibit androgen biosynthesis through targeting CYP17 may represent a viable approach in the treatment of PC.

Several studies have focused on CYP17 and indicated that inhibition of CYP17 results in inhibition of androgen biosynthesis [1] [18]. CYP17 inhibitors in the clinical trials are at various stages of drug development. The first drugs against CYP17 were designed with respect to its natural substrates (pregnenolone and progesterone) since no drugs effective on CYP17 had been reported. They are thus steroid based molecules with varying side chains [18]. Figure 2.5 demonstrates some of the CYP17 inhibitors in the clinic and clinical trials.



Ketoconazole ***Abiraterone*** ***Galeterone*** ***Orteronel***

Figure 2.5: Chemical structures of some CYP17 inhibitors in the clinic and clinical trials.

Ketoconazole has antitumor activity in PC. Ketoconazole is an antifungal imidazole agent with weak and nonspecific CYP17 inhibitory properties but although having anticancer activity in PC, it is removed from public use for this indication [19]. However, abiraterone has been overcome limitations of ketoconazole. Due to poor bioavailability of the parent molecule, abiraterone acetate was used for further examination. Abiraterone acetate research was ended up with promising results indicating that it is well tolerated, effective at lowering serum PSA level and efficient to increase the overall survival [12] . Finally, abiraterone acetate was approved by the FDA in 2011 for use in patients with metastatic CRPC both before and after docetaxel-based chemotherapy [20]. VT-464 is a small bioavailable CYP17 inhibitory molecule developed by Viamet Pharmaceuticals. It has selective inhibitory activity to 17,20-lyase reaction over 17α-hydroxylase [20]. First

studies on xenografts indicated that VT-464 has an inhibitory activity comparable to surgical castration and it is now evaluated in phase I/II study [20]. VN/124-1, also known as Galeterone or TOK-001, is a 17-benzoimidazole with inhibitory activity for both CYP17 and AR directly [14]. It has been currently licensed to Tokai Pharmaceuticals Inc. and evaluated in phase I/II trials [20]. Orteronel, TAK-700 developed by medicinal chemists at Takedea, is a non-steroidal imidazole with high specificity for CYP17 [13]. First studies on Orteronel indicated promising results that it is well tolerated and effective at lowering serum PSA level. Orteronal has been currently evaluated in phase III study [20].

Those drugs were effective in the inhibition of androgen biosynthesis; however, they possessed serious drawbacks such as poor selectivity, poor bioavailability, poor acid-stability, first-pass effects, and short half-life [21]. Although several studies have worked on the identification of drug candidates to inhibit the catalytic activity of CYP17 [1, 6, 7], a successful design of non-steroidal, specific, non-toxic CYP17 inhibitor is still an intriguing area of research. The non-steroidal compounds reported in the literature were not yet approved in clinical trials.

2.2. Drug Design Process

2.2.1. Principles

Drug candidates should be a ‘drug-like’ compound, that is, they should possess some properties indicator of its bioavailability, adequate chemical and metabolic stability, and minimal toxic effects. The five conditions, called ADMET, are the required properties for a drug candidate. ADMET is an abbreviation for absorption, distribution, metabolism, elimination, and toxicity. The drug molecule should possess membrane permeability to some extend in order to be absorbed, be orally bioavailable, therefore, should not be too

large, be target specific in order to interact with a specific protein, be catalyzed and excreted in order not to be accumulated in the body [19]. Unless all four conditions are met, the drug molecules will be unspecific or even toxic.

Since 1997, the Lipinski rule of five is used to estimate the drug-likeliness of a candidate molecule [21] by its chemical composition. According to Lipinski et al., a drug-like molecule should have:

- not more than 5 hydrogen bond donors
- not more than 10 hydrogen bond acceptors
- a molecular weight under 500 g/mol
- a partition coefficient logP less than 5

Lipinski rules define the properties required for a molecule to be considered as drug candidate but not adequate to identify a molecule as a potential drug. Binding affinity is also a major factor in computational drug design, which is correlated with the overall energetic favorability of the drug binding. Affinity is associated with the enthalpy change (ΔH) and the entropy change (ΔS) in the binding process, and is directly related with the Gibbs free energy of binding (ΔG). The relation is:

$$\Delta G = \Delta H - T\Delta S, \text{ where T refers to temperature.}$$

For a reaction to be favorable, ΔG needs to be minimized and, in most cases, enthalpy minimum is preferred over entropy maximum [19]. Maximized entropy results in flexibility, but a limit for flexibility should be set in order to protect stability of the drug molecule. Stability of the drug candidate can be satisfied with the addition of side chain atoms to the drug molecule with the consideration of Lipinski rule of five. The balance between binding affinity and the Lipinski rules should be kept since both are essential

requirements in drug development. Enhancing the binding affinity without a disruption in the Lipinski rules is thus an important step for a successful drug design.

2.2.2. Computational Drug Design

There are a wide range of approaches ranging from experimental methods and combinations of computational methods with experimental methods. High throughput screening (HTS) is one of the purely experimental methods used to test a large number of small molecules against a specific protein. Using robotics and chemistry, chemical or pharmacological tests for millions of small ligands and their inhibitory activity can be conducted quickly [8]. The results of these experiments provide starting points, called lead compounds or hits, for drug design. However, a high experimental cost and the small numbers of available ligands are the major drawbacks of purely experimental techniques in general [8].

X-ray crystallography and NMR technologies used in determination of protein structures initiate a new approach in drug development, called computational or structure-based drug design. Computational drug design has been accepted and applied over the years in both industry and academia. Using protein structure data, the active sites of proteins can be investigated and small molecules acting on these active sites can be design and tested in silico. Instead of screening millions of small ligands experimentally, first drug screening can be performed by computational tools like Gold, Molegro, Autodock to decrease the number of ligands to be screened experimentally. These computational drug screening tools estimates binding energies (scores) of the ligands onto the target protein by calculating inter- and intra- atomic forces. Once compounds with the highest scores or the best binding energies are found, bioactivity of these molecules can be tested experimentally. By this end, number of ligands to be screened against target protein can be

reduced; therefore, the high experimental cost required for HTS can be reduced with combination of computational and experimental techniques. Once a candidate (lead) with inhibitory activity in micromolar quantities is discovered, lead optimization process is started to increase inhibitory activity to the nanomolar quantities. The lead optimization and lead discovery processes are repeated several times to identify better drug candidates. Thus drug design is an iterative process as summarized in Figure 2.6. This iterative approach is widely used now and drugs developed with this approach include drugs against AIDS, cancer, hypertension. This study focused on lead optimization part of the process to develop a lead compound with a fragment based drug design approach in order to increase its binding affinity and inhibitory activity.

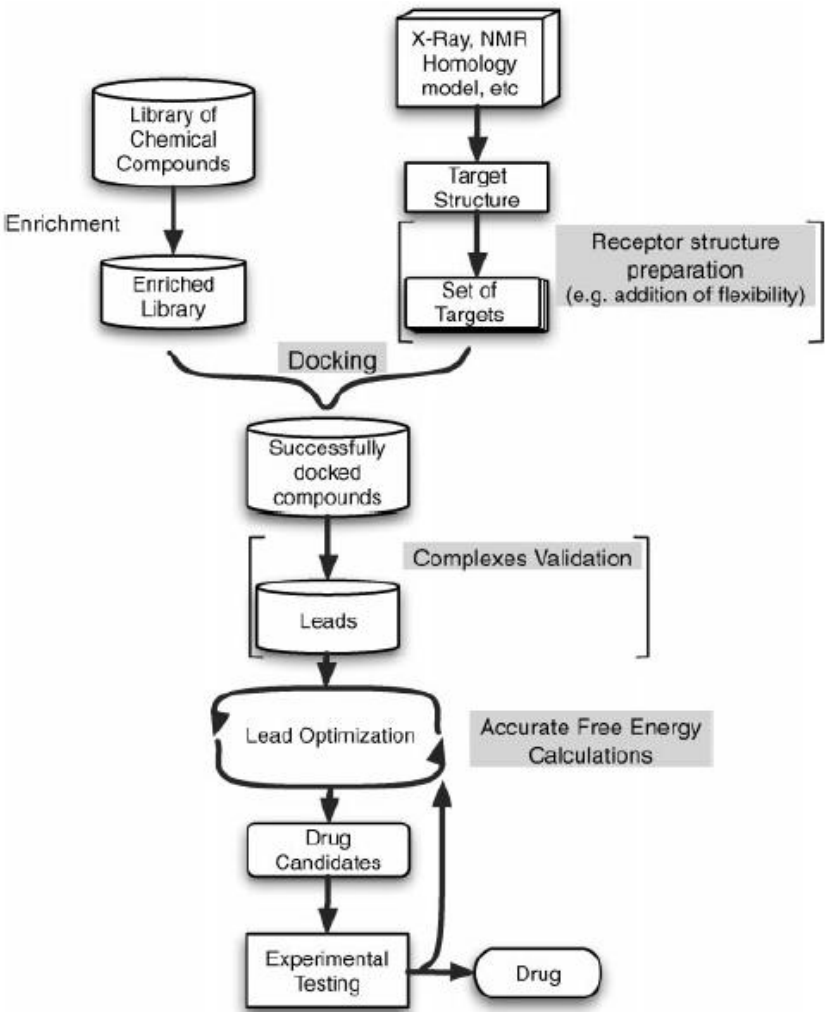


Figure 2.6: Common process in computer-based drug design research [22]

2.2.3. Fragment-Based Drug Design

Fragment based drug design (FFDD) is a tool in combinatorial chemistry that is developed recently and has the advantage of increasing productivity in drug design. In this approach, instead of considering whole molecule, the optimization of binding affinity is calculated as the sum of the individual fragment interactions. Once small chemical

fragments binding to the biological target are identified, even though weak binding affinity they have, they can modified by adding new side chains or combining with other lower affinity fragments by FBDD approach to produce a lead with a higher affinity.

Developing libraries of fragments is the first step of the fragment-based drug discovery. There are some unique considerations for fragment libraries. For example, fragment should be smaller than typical HTS compounds. The fragments should be low-molecular-weight ligands (<250-300 Da) since they will be expanded or linked together in the production of high affinity and high potency drug leads [23]. Besides, instead of Lipinski rule of five, rule of three, which is molecular weight < 300 Da, cLogP<3, the number of hydrogen bond donors and acceptors each < 3 and the number of rotatable bonds < 3, is the mainstay of FBDD [24].

Detection of bound fragments with an even weak interaction to the protein target and characterization of their binding properties requires very sensitive methods like differential screening fluorimetry (DSF), nuclear magnetic resonance (NMR), X-ray crystallography, isothermal titration calorimetry (ITC). These methods are associated with one or another drawback such as low throughput, occurrence of false positives or false negatives, high sample consumption and immobilization of proteins [25]. Therefore, combinations of several methods should be used for a reliable detection and characterization of ligand binding. Ciulli et al. represented a three-stage screening methodology which is composed of DSF for preliminary screening, NMR for hit validation, ITC and X-ray crystallography for binding characterization [26]. FBDD is a new but a successful tool in drug discovery and its success can be improved by the new technologies and new approaches. The first successful implementation of FBDD is the drug Vemurafenib approved recently by the FDA in 2011 for the treatment of metastatic melanoma [27].

FBDD has been recently used in in silico drug design as the lead optimization process. Examples of several software packages with de novo design algorithms that rely on different scoring algorithms are listed in Table 2.1. These algorithms identify a ligand-protein interaction pocket and construct new compounds by combinatorial or sequential assembly of molecular fragments [28]. These methods use sets of pre-defined fragments and constructs new structures from an anchor fragment (scaffold) by connecting side-chains (fragments) to it via a set of linkers.

Table 2.1: Examples of de novo drug design algorithms [a]

Method	Concept
BUILDER	Recombination of docked molecules, combinatorial search
CONCERTS	Fragment-based, stochastic search
HOOK	Linker search for fragments placed by MCSS
LUDI	Fragment-based, combinatorial search
SPLICE	Recombination of ligands retrieved by a 3D database search

[a] Additional examples can be found in Ref. [28].

2.3. Current Approach

2.3.1. Nonlinear Programming

In experimental science, some data analysis can be linear problems, but in general most of them are non-linear problems. Nonlinear programming (NLP) is the process of solving an optimization problem with an objective function and some constraints, where either one of them is nonlinear [29]. NLP solves the problem in such a way to optimize (maximize or minimize) a given objective function. Constraints are the sum of equalities or inequalities in the system defined. There is no guarantee that any solution exists that satisfies all of the given constraints. Therefore, data fitting problems are applied to find a solution for the nonlinear functions that best fits the given constraints.

In order to find the better estimates for the constraints, the methods used in data fitting problems include the least square regression method. Least square regression is a widely used and efficient model in classification problems. Least square method works as summation of the squares of the experimental deviations (errors made) in the solutions of every single equation [29]. Optimal values of the parameters for the given constraints can be calculated in the least square sense since the least square method is able to minimize a quantity with minimum error rates by giving a best-fit curve.

2.3.2. Proposed Approach

The aim of the present study was to create novel CYP17 inhibitors with potential to be developed into potent, selective and orally bioavailable compounds by targeting CYP17 heme catalytic region. A new fragment-based drug design approach was applied to develop previously identified lead compound. To develop the lead and increase its binding affinity,

fragment sites on the lead were determined and contributions of different fragment combinations on ligand binding were studied. The study was assumed to be a NLP problem since fragments may possess a combinatorial effect on binding energy. Objective of this approach is to minimize binding energy and thus the least-squares is used to minimize quantity with minimum error rates. Flowchart of the model used in this thesis in shown in Figure 2.7.

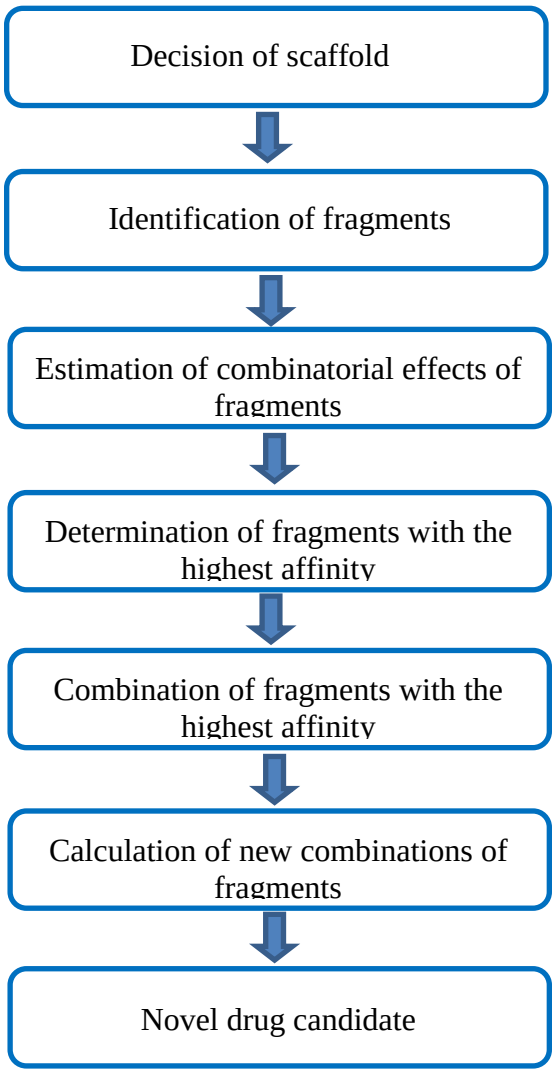


Figure 2.7: Flowchart of the Model used in this thesis

Chapter 3

Materials and Methods

3.1 Molecular Dynamics Simulation

Common experimental methods (i.e. NMR, X-ray crystallography, et al) only can provide time-averaged structures of molecules. However, rather than being relatively rigid structures biomolecules are highly dynamic systems in which the internal motions and resulting conformational changes play an essential role in their function [30]. The paradigm structure-function has been shifted towards the structure-dynamics-function triad, where not only the knowledge of the tertiary or quaternary structures of biomolecules is important to understand their functions, but also the dynamic motions causing conformational changes on different timescales [31]. These small conformational changes over time can resemble functionally important states caused by modifications, mutations, and interactions with other biomolecules thus having an impact on the related biological mechanism. In order to have an idea of the movements of atoms in a biomolecule microscopically, especially to study the functionally-dependent conformational transitions responsible for i.e. stability and binding between biomolecules, simulations under theoretical models are necessary. In molecular dynamics (MD) simulation, the motion of each atom and the conformational changes in biomolecules can be simulated as a function of time by numerical integration of Newton's equation of motion [31] and the link between structure, dynamics, and function can be generated in silico.

MD simulations applied to biological sciences are important tools for investigating the physical basis of the structure and function of biological molecules to integrate experimental studies. With the advances in MD studies and developing technologies, effects of solvents and temperature on structures of larger proteins can be studied too. The first published MD simulation study was applied to the bovine pancreatic trypsin inhibitor by McCammon et al. in 1977 and that was a short and in vacuum simulation [32]. Thirty-eight years later, MD simulations are widespread tools used to study even larger macromolecules with surrounding solvent and ions in the ~1000 times longer simulations, where even experimentalists use to rationalize experiments. Today main goals of MD simulation studies in structural biology include understanding the driving forces for protein folding, prediction of protein folding and association, and understanding the motions of biomolecule and how this is coupled to its function.

The most common tools for MD simulations available to studies in life sciences include Amber [33], CHARMM [34], GROMOS [35], and NAMD [36]. Among them, NAMD with CHARMM force field parameters was used in this thesis. VMD [37] is a graphical user interface (GUI) used to prepare and evaluate MD simulations.

3.1.1 NAMD

NAMD is one of the MD simulation tools that enable high-performance simulation of large macromolecular systems in a realistic biological environment. NAMD requires at least four files; a protein data bank (pdb) file, a protein structure file (psf), a force field parameter file, and a configuration file. The pdb file, which stores atomic coordinates and/or velocities for the system, can be obtained via the Internet at http://www.pdb.org. The psf, which stores the structural information of the protein, can be generated via the GUI tool VMD. The force field parameter file, which stores a mathematical expression of

the system potential, can be generated by Quantum Mechanics (QM) or can be found online. Force field parameters that can be used in NAMD simulations include AMBER, CHARMM, X-PLOR, and GROMACS. The configuration file should contain user-specified options that NAMD should adopt in running simulation.

In MD simulations, atomic trajectories (position and velocities) in the system is determined solving the Newton's equation of motion iteratively for each atom in the system, the interactions between atoms are described using molecular mechanics force fields, temperature and pressure are controlled using statistical mechanics methods, and the electrostatic forces are evaluated using partial mesh Ewald (PME) [36]. In these simulations the atoms move according to the Newton's equation of motion:

$$\vec{F_i} = m_i \vec{a_i} \tag{3.1}$$

where $\vec{F_i}$ is the force, m_i is the mass, $\vec{a_i}$ is the acceleration of atom i. NAMD uses force field parameters defined at the beginning of the simulation to calculate total potential energy acting on every atom due to inter- and intra-atomic forces.

Total potential energy is calculated as the sum of the bonded and non-bonded interaction potentials:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb}. \tag{3.2}$$

The first three terms define the bonded interactions of stretching, bending, and torsional interactions:

$$U_{bond} = \sum_{bond\ i} k_i^{bond} (r_i - r_{0i})^2 ; \tag{3.3}$$

$$U_{angle} = \sum_{angle\ i} k_i^{angle} (\theta_i - \theta_{0i})^2 ; \quad (3.4)$$

$$U_{dihedral} = \sum_{dihedral\ i} k_i^{dihe} [1 + \cos(n_i \varphi_i - \gamma_i)] \quad (3.5)$$

where bonds defines bond stretching counting each covalent bond in the system, angles defines bond bending counting the angles between each crossing bonds, and dihedrals defines bond torsions counting atom pairs separated by exactly three covalent bonds with the central bond subject to the torsion angle φ. The final two terms in 3.2 define the non-bonded interactions of Van der Waal's (vdW) forces and electrostatic interaction potentials:

$$U_{vdW} = \sum_{i} \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \quad (3.6)$$

$$U_{Coulomb} = \sum_{i} \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \quad (3.7)$$

where vdW defines an approximation by Lennard-Jones 6-12 potential and Coulomb defines the electrostatic interactions. [36]

Once total potential energy is calculated, the force on each atom can be computed from a change in potential energy between current and new positions:

$$\vec{F_i} = -\frac{\partial U}{\partial r_i} \quad (3.8)$$

where r is the distance and U is the potential energy. Given the knowledge of potential energy (calculated in 3.2) and initial coordinates (given in initial psf), force acting on each

atom can be calculated. Atomic forces and masses are then used to determine atomic positions over series of very small time steps:

$$\vec{F_i} = m_i \frac{\partial^2 r_i}{\partial t^2} \quad (3.9)$$

where t is the time. As new atomic positions are determined, the calculations are started over with the new coordinates. This provides a trajectory of changes in atomic positions over time. [38]

3.1.2 NAMD Setup

MD simulations were carried out on Koç University High Performance Computing (HPC) Laboratory in parallel; using NAMD 2.6 [36] and two force field parameters: CHARMM27 [39] and custom_heme [40]. Coordinate of the initial structure was obtained from Protein Data Bank (PDB code 2c17) [41] which is CYP17 (Steroid 17-alpha-hydroxylase/17,20 lyase). VMD was used to prepare psf file for NAMD simulation. Using psfgen package, atom and residue names were replaced with ones recognized by NAMD. Protein and heme moiety is initially treated as separate chains, and linked after psfgen made necessary replacements. The topology file specific to heme residue and parameters for heme-cystine link are set according to the work of Autenrieth et al. [40]. The protein was solvated in a water-box with a min 10 Å distance from any atom in the protein to the boundary by VMD Solvate plug-in. Counter ions were added to neutralize the system by VMD Autoionize plug-in.

At first the system was minimized through 10000 steps by fixing the backbone and heme, and then fixed atoms were released through 20000 steps. Second, the system was brought to physiological temperature (310 K) with 10K increments (10ps simulation at each temperature increment). After the minimization step, molecular dynamics simulation

was carried out with constant temperature (310K) and pressure control using Langevin piston method.

The time-step of the simulation was set to be 2 fs and, the bonded interactions, the van der Waals interactions (12 Å cut-off), and the long-range electrostatic interactions with partical-mesh Ewald (PME) were included in the calculations to define the forces acting on the system. The damping coefficient was set to be 5 ps-1 using Langevin dynamics to handle pressure control. Constant pressure was held at 1atm with decay period of 100fs and a damping time of 50fs. The simulation was carried out for 10 ns. Related configuration files can be found in Supplementary documents.

3.2 Protein-Small Molecule Docking

Docking techniques have been routinely used in drug-design process: from hit identification to lead optimization stages and beyond [22, 38]. It is a computational procedure that attempts to predict noncovalent binding of a small molecule (ligand) and a macromolecule such as a protein efficiently. Main goal of protein-small molecule docking studies is to determine the bound conformation and the binding affinity of a ligand to a macromolecule, starting with their unbound structures [42].

Given a resolved or modelled structure of a target protein, docking simulations with the quick screening of large libraries of small molecules in silico are performed primarily in order to identify putative hit compounds. Several scoring functions are designed to predict the binding affinity through evaluation of interactions between ligands and targets [38, 42]. Top scoring ligands after docking simulations are identified as hits for later process. Later on drug design process, an accurate prediction of the bound conformation of the hits to target proteins is of particular practical importance for the lead optimization. With the accurate prediction of binding conformations, modifications to known active

structures in order to increase their binding affinity can be performed efficiently according to binding pocket and interacting sites of the target protein and the hit [38] and can be tested in silico before compound synthesis. Therefore, by docking studies, the optimum conformation (best-fit) that maximizes the interaction between ligand and the target is screened.

In general, docking studies are composed of two main components: a search algorithm to find possible conformations of the protein-ligand complex and a scoring function to estimate relative binding affinities and to score alternative poses of the ligands. The most common search algorithms involve Monte Carlo, simulated annealing, and genetic algorithms (GA) [42]. Scoring functions used today can be classified as either force field based, empirical, or knowledge based scoring functions [43]. Force field based scoring functions describe binding affinity as the summation of both intra- and intermolecular interactions in the system. Empirical scoring functions describe the terms accounting for van der Waals interaction, hydrogen bonding, deformation penalty and hydrophobic effect [44], whereas knowledge based scoring functions calculate pair potentials from statistical observations of intermolecular contacts [45].

Several docking tools using various combinations of search algorithms and scoring functions have been developed to date, including DOCK, GOLD, Flex and AutoDock. These methods introduce different approximations to simplify the complexity of docking – for example, the rigid-body assumption. Although default values are set in order to dock rigid ligands into rigid receptors, with the advances in in docking algorithms, ligand flexibility and, to less extent, protein mobility can be incorporated into the docking experiments [22]. Rigid-body assumption, in which the protein structure is fixed, can save computational time with the pre-calculation of forces acting on the ligand, which is advantageous for screenings of large databases.

AutoDock, which is freely available docking tools, was used for protein-small ligand docking in this thesis. The version AutoDock 4.2, which will be addressed as AD4 later in this thesis, is preferred over Vina, the final version of AutoDock, since AD4 can compute both binding and docking energies unlike Vina computing only binding energy.

3.2.1 AutoDock 4

AD4 predicts the optimal conformation of protein-ligand complex and binding affinity of the ligand to the protein by the assumption of rigid-protein and flexible-ligand structures [46]. AD4 offers several alternatives of search algorithms such as Monte Carlo Simulated Annealing, Genetic Algorithm (GA) and Lamarckian-Genetic Algorithm (LGA). LGA search parameter was used in this thesis as LGA is suggested by the producers for the best in high-degrees of freedom [46].

LGA is an iterative search algorithm in which individual conformations search for their local minima and then pass this information to later generations till the end of number of runs defined. In LGA searching, a random population of trial conformations is generated at first, and then, in successive generations, these conformations are subjected to genetic operators like mutations and crossovers and compete in an evolutionary manner, ultimately individual conformations with lowest binding energy is selected. A force field based scoring function is used to predict binding energies of small molecules to their targets. By evaluating energies for both the bound and unbound states, this force filed based scoring function incorporates intramolecular energies into the predicted free energy of binding [46].

A grid-based approach used in AD4 calculates a special atomic affinity grid maps for each type of atom in the ligand, where every atom is assigned a non-bended interaction potential with the protein and electrostatic potentials. A Gaussian function is constructed

with zero energy at the site of attachment and steep energetic penalties at surrounding areas [46]. The docking analysis is then performed by assigning a special atom type in the ligand for the atom that forms the covalent linkage. At the end of docking simulation, AD4 computed three energies as outputs: intermolecular energy, internal energy, torsional free energy, and unbound system's energy. The sum of the first two energies accounts for the docking energy, while the sum of the first, the third and the fourth energies accounts for the binding energy.

3.2.2 ADT4 Setup

AutoDock requires 4 input files for docking simulations: a PDBQT file for both macromolecule and ligand molecule, Grid Parameter File (GPF) for the grid parameters to be docked into, and Docking Parameter File (DPF) for the docking parameters. GUI platform of AutoDock Tools (ADT) can prepare all the necessary files for docking.
PDBQT files include the information needed by both AutoGrid and AutoDock, which involves polar hydrogen atoms, partial charges, atom types, and information on the articulation of flexible molecules. PDBQT files are prepared for the protein and the ligand separately. In dockings where selected amino acids in the receptor are treated as flexible, a third file that includes the coordinates of the atoms in the flexible portions of the receptor should be prepared.

Preparation of PDBQT files involves: adding hydrogen atoms to the molecule, adding partial charges, deleting non-polar hydrogens and merging their charges with the carbon atoms, assigning atom types as hydrogen bond acceptors and donors and aromatic and aliphatic carbon atoms, choosing a root atom that will act as the root for the torsion tree description of flexibility, and defining rotatable bonds and building the torsion tree. In preparation of PDBQT file for the protein, there was no need to add polar hydrogens to

CYP17 since the protein already contains polar hydrogens after MD simulation. Preparation of PDBQT files can be done by either scripting or through GUI representation of AD4.

Grid maps contain the information for a three dimensional lattice of regularly spaced points, surrounding and centered on some region of interest of the macromolecule. Grid maps, one for each atom types present in the ligand being docked, were calculated by AutoGrid with respect to grid parameter file, GPF, specified. The GPF can be prepared by ADT and defines a grid box to specify the protein region to be docked into. Therefore the GPF file contains the grid center coordinates of this box, the grid size as the number of points, the spacing between two grid points and types of the atoms present in the ligands to be mapped. For this study, center of the grid box was defined as the center of the macromolecule, map size was defined as 52x52x52, and spacing was kept as default 0.375 Å. These parameters were determined to cover the entire active site of the protein. The active site of the protein was determined with respect to crystal structure of CYP17 in complex with its inhibitor Abiraterone to cover the region that Abiraterone targets [47].

DPF stores the parameters for docking algorithm and the options for the method. The defined algorithm was the Lamarckian-Genetic Algorithm and the options were the default values where number of runs was 10, population size was 150, maximum number of evaluations was 2500000, number of generation was 27000, crossover rate was 0.8, mutation rate was 0.02, and crossover rate was 0.8.

3.3 Optimization Problems

Optimization problem is identified as the problem of finding the best solution among all feasible solutions. In a simplest case, the problem is to optimize (minimizing or maximizing) a real function by choosing the input within an allowed set of values and

computing the value of the function. And in a general definition, an optimization problem seeks for the best available values of some objective function given a set of constraints.

The type of objective function and/or constraints determines the approach to be used in the optimization model. Several examples of optimization problems include linear programming in which both the objective function and the constraints are linear, stochastic programming in which some of the constraints of parameters depend on random variables, and nonlinear programming in which the objective function and/or the constraints contain nonlinear parts.

Once the problem is determined, and the objective function and set of constraints are defined in a suitable programming language, the optimization software can easily solve the problem in a short time. There is variety of optimization software, including LINCOA, MATLAB Toolbox, and General Algebraic Modeling System (GAMS). Among them, GAMS is used to solve the proposed model in this thesis.

3.3.1 GAMS

GAMS, a high-level modeling system for mathematical programming and optimization, is specifically designed for modelling linear, nonlinear and mixed-integer optimization problems. It is tailored for complex, large scale modeling applications, and allows building large maintainable models that can be adapted quickly to new situations. It is and user-friendly tool and easy to manipulate formulations defined or to change from one solver to another.

3.3.2 GAMS Setup

Initial structure of scaffold is designed with respect to the lead compound identified by Armutlu et. al. The lead compound, shown in Figure 3.1, was determined by a structure-based drug design approach, in which a large library of small molecules screened in silico and top ranking molecules were screened experimentally [9].

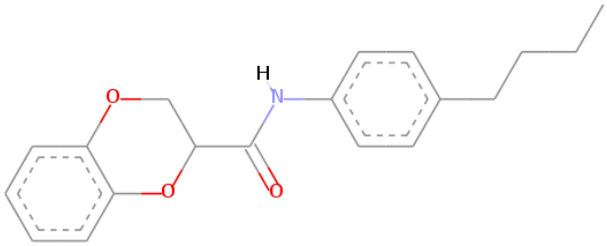


Figure 3.1: Structure of the lead discovered by Armutlu et al [9].

The scaffold for this study was determined from the binding position of the lead in order to develop new compounds with a higher affinity than the lead. Four positions to insert fragments, R1, R2, R3 and R4, were decided with respect to chemical interactions between ligand and protein as indicated in Figure 3.2. R1 position is for an aromatic group to interact with the heme residue on the target protein, R2 position is donated to an electron donor group for coordination to the iron atom of heme, R3 position is for an electron donor too to increase electronegativity, and R4 position is allocated for an alkyl group for hydrophobic tail.

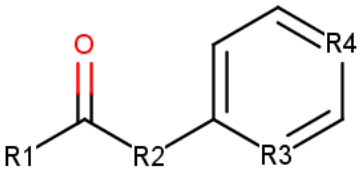
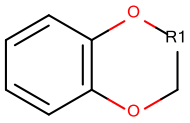
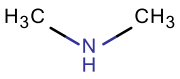
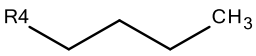
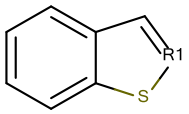
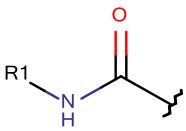
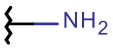
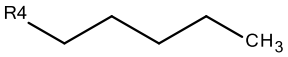
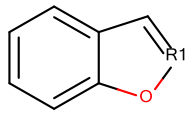
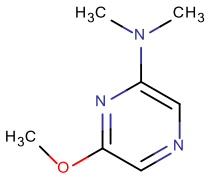
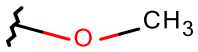
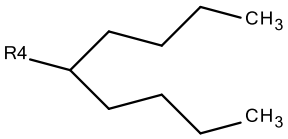
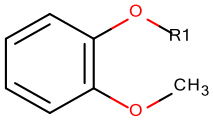
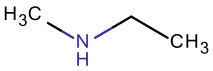
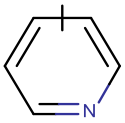
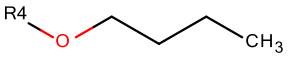
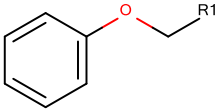
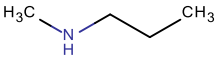
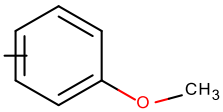
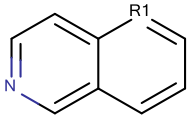
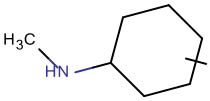
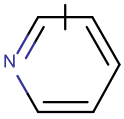
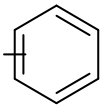
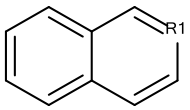
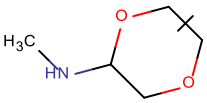
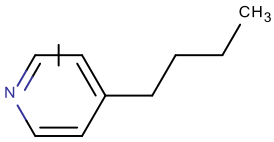
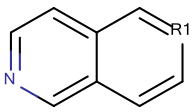
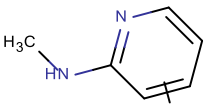
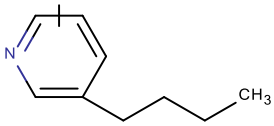
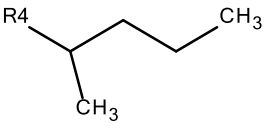


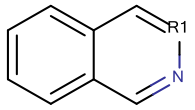
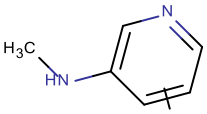
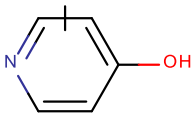
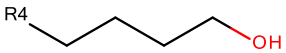
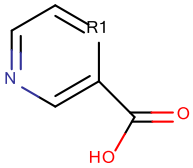
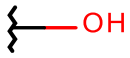
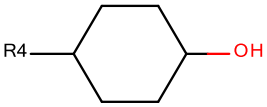
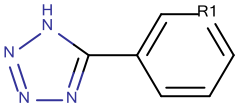
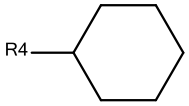
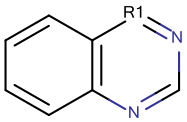
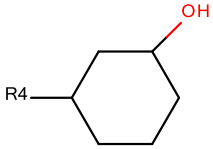
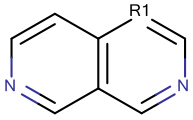
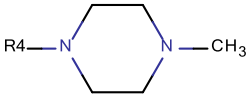
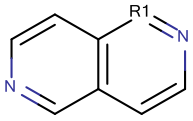
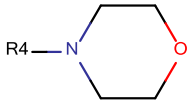
Figure 3.2: The Scaffold structure to be used as a template to develop novel drugs with a fragment-based drug design approach.

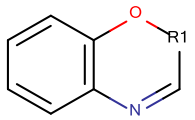
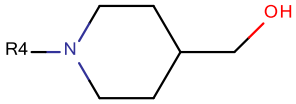
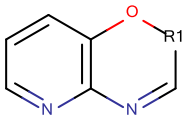
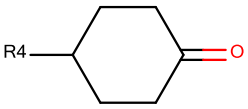
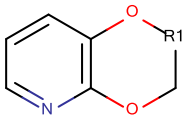
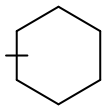
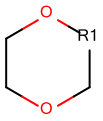
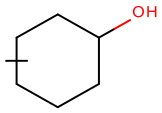
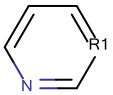
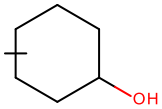
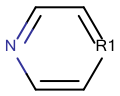
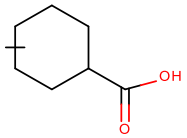
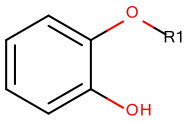
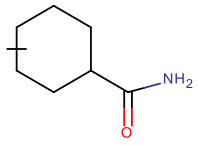
Fragments to be inserted into the scaffold were determined initially from the compounds synthesized from the Acar's Research Group at Boğaziçi University and extended to increase diversity. At the end, the total number of fragments with respect to binding positions was determined as R1:30, R2:9, R3:10 and R4:31, given in Table 3.1.

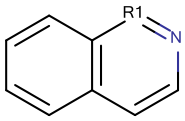
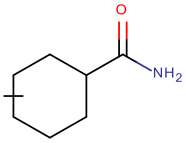
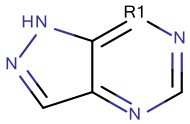
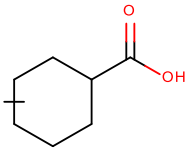
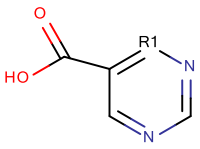
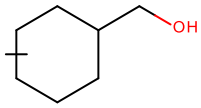
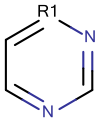
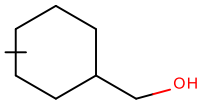
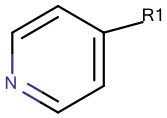
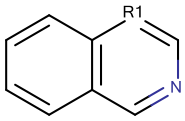
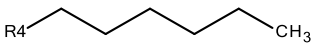
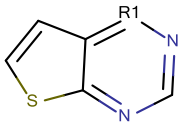
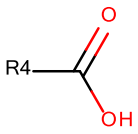
Table 3.1: List of fragments to be combined in the discovery of novel molecules

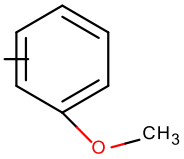
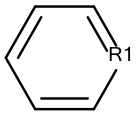
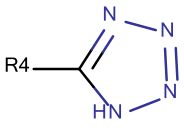
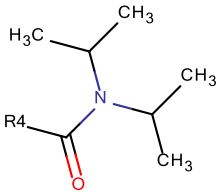
	R1	**R2**	**R3**	**R4**
1				
2		(reverse) 		

3				
4				
5			--N	
6				
7				
8				

9				
10				
11				
12				
13				
14				

15				
16				
17				
18				
19				
20				
21				

22				
23				
24				
25				
26				--N
27				
28				

29				
30				
31				

New molecules designed randomly by the combinations of fragments determined. New molecules were generated with the combination of fragments and drawn by either Spartan or Discovery Studio computational protein visualization tools. In total, 621 molecules were designed randomly. Binding and docking energies were calculated by AutoDock 4.

A nonlinear programming model was generated by GAMS software to calculate combinatorial effect of fragment combinations. Based on constraints of fragment combinations and respective calculated energies, contributions of every fragment into calculated energies were estimated by the model generated. The objective function of this model was defined in the least square sense to minimize binding and docking energies. The nonlinear programming model used in this thesis is described by the following expressions:

$$\text{Min } z = \sqrt{{r_1}^2 + {r_2}^2}$$

$$\text{s. t.}$$

$$DE_i = \sum_j x_j t(i,j) + r_{1i} \forall i$$

$$BE_i = \sum_j x_j t(i,j) + r_{2i} \forall i$$

$$x_j, y_j \in R^n$$

$$r_{1i}, r_{2i} \in R^n$$

$$t(i,j) \in \{0,1\}$$

where DE_i is the docking energy of molecule i, BE_i is the binding energy of molecule i, r_{1i} and r_{2i} are the estimation errors for molecule i, x_j and y_j are weight coefficients of fragment j into binding and docking energies respectively, and t(i,j) is the binary variable table of the j^{th} fragment of molecule i. t(i,j) binary variable table was designed with respect to existence of indicated fragment j in the structure of molecule i. If the fragment exists, the corresponding value in t(i,j) is 1, otherwise, it is 0. The model was created and solved by GAMS software.

GAMS gave the best solution minimizing least square of summation of estimation errors for both binding and docking energies. Since there were two objective functions as binding and docking energies to be minimized, the problem is multi-objective optimization problem. Unfortunately, there may not be a single solution for multi-objective optimization problems. That's why; the original problem with multiple objectives was converted into a single-objective optimization problem by minimizing the total estimation errors of binding and docking energies.

The model estimated weight coefficients, x_j and y_j, of every fragment j for both binding and docking energies by minimizing the total energy. With the calculation of weight coefficients, x_j and y_j, fragments having the minimal contribution into total energy were identified. Combinatorial effects of different fragments were analyzed. Combinations of fragments with optimal contribution to total energy were predicted and used in generation of new molecules with lowest binding and docking energies.

Chapter 4

Results and Discussions

4.1 Analysis of Cys-Heme link

The usual topology files (CHARMM27 etc.) does not contain necessary information for HEME residue. That's why, several other researchers studying CYP proteins have focused on patches specific for HEME residue. Unfortunately, there are still limited resources on how to link HEME residue to a CYS (cysteine) residue. Topology and parameter files specific for HEME residue that was generated by Autnrieth et al., were used in this study. Pdb file was modified and psf file were generated using VMD and its plug-ins. The final structure of successfully linked Fe-S bond was shown via PyMOL version 1.6.9.0 in Figure 4.1.

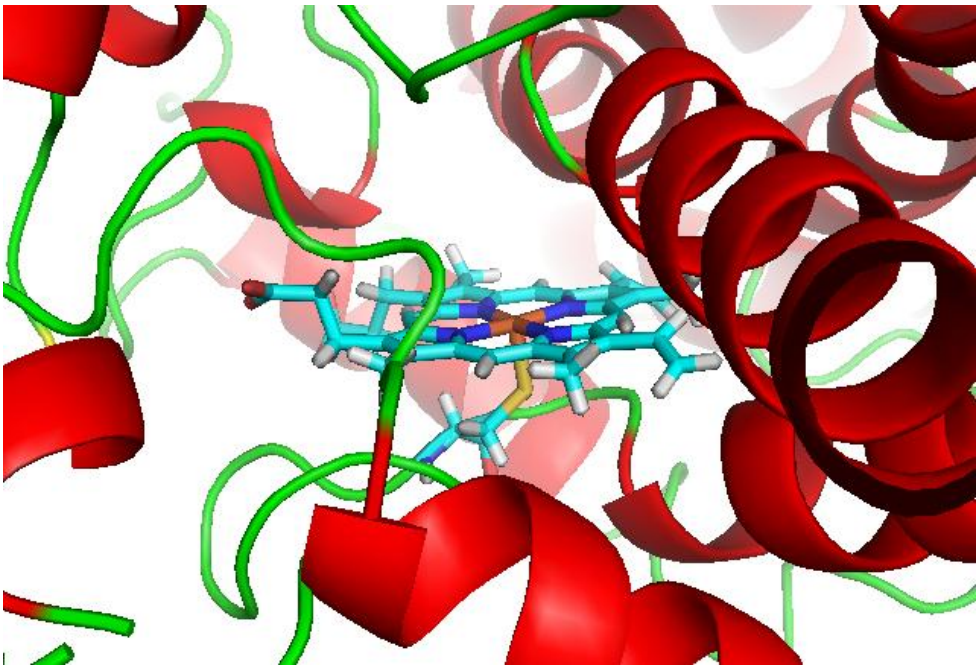


Figure 4.1: Visualization of successfully linked Fe-S bond via PyMOL. Protein was colored by secondary structure where HEME and Cys395 were colored by atom types. The HEME residue was centered and focused to get a better view of the Fe-S bond.

4.2 Analysis of Molecular Dynamics Simulation

Molecular Dynamics simulation was applied for 10 ns and a stable CYP17 structure at physiological conditions was observed through end of the simulation. Root-mean square deviation (RMSD) plot was obtained for each frame by aligning the final structure of CYP17 at the corresponding frame to the initial structure. MD simulation analysis was performed on VMD and its plug-ins.

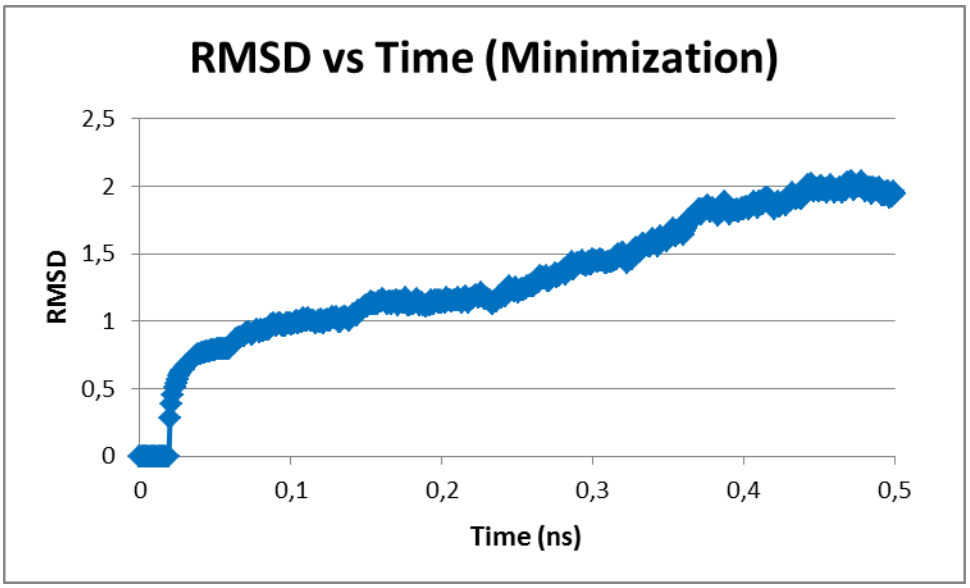


Figure 4.2: RMSD vs Time Plot for Minimization

RMSD vs Time graph for the minimization part of the simulation was shown in figure 4.2. The graph shows an expected pattern of increase: at the first 20ps of the simulation, which is the minimization with fixed backbone and heme, only hydrogens atoms were allowed to coordinate themselves in a minimum energy conformation, then the fixed atoms were allowed to coordinate themselves for 40ps without pressure control. Finally the whole system was heated up gradually to physiological temperature (310 K) and equilibration step was started. CYP17 showed an expected pattern of solvation and relaxation during minimization stage and no deformations were observed. Figure 4.3, including both minimization and equilibration steps, indicates RMSD trajectory plot for CYP17 through total 10 ns of MD simulation.

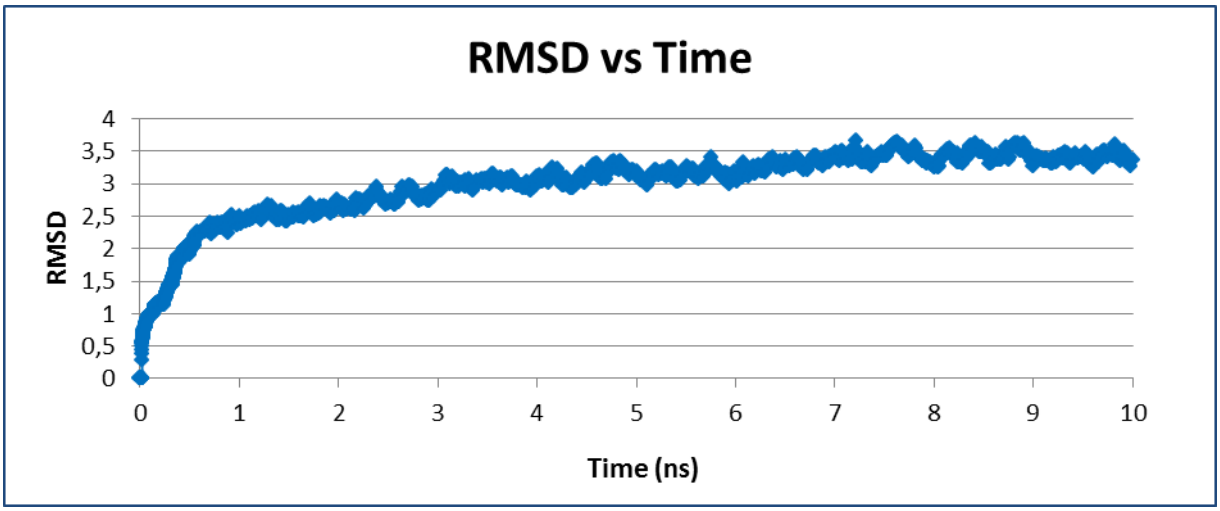


Figure 4.3: RMSD vs Time graph for both minimization and equilibration

As shown in figure 4.3, RMSD increased gradually at first 0.7 ns simulation, including minimization, up to 2,5 and continued to increase through 6 ns of simulation. After 7 ns of simulation, there were not any significant deviations observed during the rest of the simulation and RMSD value converged to ~3,5 Å. The convergence of RMSD indicates the stability of the given structure in the simulation. Also this behavior of convergence, indicating stability, is a significant property for research in computational drug design since instability and deformations in the given structure might hinder a reliable rigid binding pocket targeted by docking studies.

4.3 Validation of Autodock with Known Drugs and Natural Substrates

Known drugs and natural substrates of CYP17 were screened virtually by AutoDock before discovery of novel molecules. Calculated binding and docking energies, shown in table 4.1, were later used to validate and compare novel molecules identified in this thesis.

Table 4.1: Binding and docking energies of natural substrates (progesterone and pregnenolone) and known drugs (Abiraterone and Ketoconazole) of CYP17, calculated by AutoDock 4.

ID	**Binding Energy** (kcal/mol)	**Docking Energy** (kcal/mol)
Progesterone	-10,91	-11,39
Pregnenolone	-10,37	-11,10
Abiraterone	-12,36	-13,14
Ketoconazole	-7,73	-11,08

Both binding and docking energies of substrates are around -10 - -11 kcal/mol. Those results were used as an indicator of strong affinity towards CYP17. Abiraterone has the minimum binding and docking energies around -12 - -13 kcal/mol which is the target value to develop new molecules in this thesis. Ketoconazole has docking energy -11 kcal/mol similar with substrates but lower binding energy which might be an indicator of its non-specific binding.

Binding conformation of both substrates and known drugs can be seen in Figure 4.4. Active oxygen or nitrogen atoms in the molecules were coordinated towards heme for its catalytic activation. This property was also taken into account while designing new molecules to be used in the proposed model.

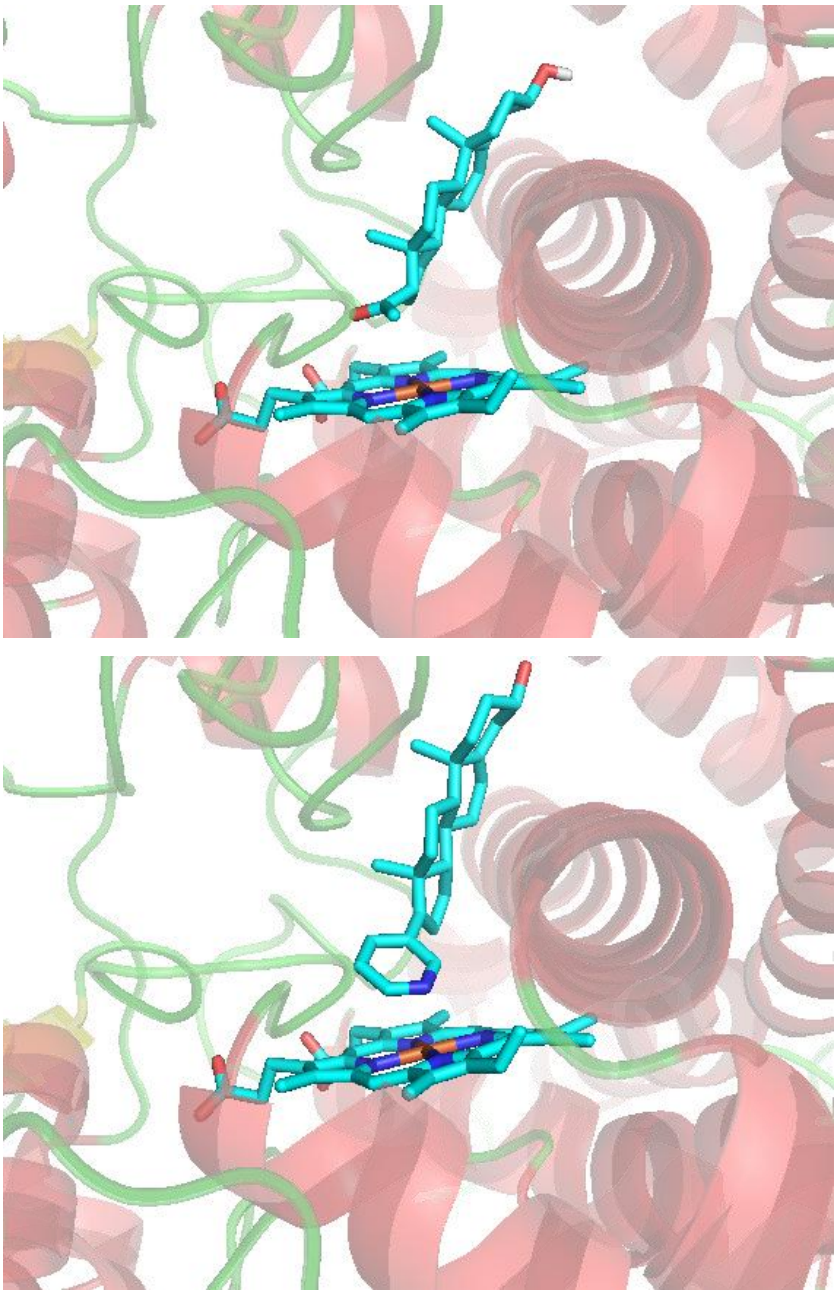


Figure 4.4: Binding conformations of a natural substrate, pregnenolone, and a known drug, abiraterone, onto the protein, CYP17. Pregnenolone binding is shown at the top and ketoconazole binding is shown at the bottom.

4.4 the Two Stages of the Proposed Model

The proposed model was composed of two stages. Given the binding and docking energies, contributions of each fragment into energies were calculated as the first stage of the model. Objective function of the model was designed to minimize summation of squares of estimation errors for binding and docking energies. The aim is to identify small molecules with best binding energies and also best docking energies. Since binding and docking energies are not always increase or decrease in a synergy, the aim is to minimize their summation. The constraints of the model involve weight coefficient for each fragment, existence parameter of each fragments and estimation errors for binding and docking energies. The equation to solve by the algorithms was shown in equations 3.11-12. Once determining the best fragments that have the highest contribution to the total energy, designing of the new small molecules is the second stage of the model. New combinations of fragments were generated in order to identify top ranking small molecules composed of those predetermined fragments shown in Table 3.1.

The purpose of this study is to identify the top ranking molecules that can possible be generated from some predetermined fragments given a small subset of ligand library generated by the combination of those fragments. The proposed approach was verified first with a small library of small molecules that all possible combinations of fragments were known.

4.5 Verification of Proposed Approach

As a first step, a small verification library having the same scaffold and fragment positions with R1:5, R2:2, R3:5, R4:3 fragments were designed. At total, verification library was composed of 150 molecules including all possible combinations of limited

number of fragments. For fragment site R1, fragments 1st, 17th, 22th, 29th, and 30th; for R2, fragments 1st and 5th; for R3, fragments 1st, 4th, 7th, 15th and 17th; and for R4, fragments 1st, 4th and 5th, were chosen to build a verification library in order to test accuracy of the proposed methodology.

All molecules were drawn by Discovery Studio and were screened in silico by AutoDock 4 docking tool. 60 of designed molecules were chosen randomly, and contributions of each individual fragments into energies were calculated using GAMS. Besides small (~6%) deviations in the estimated total binding and docking energies with respect to calculated energies, the model was successful in the identification of the top ranking molecules. The model can predict top 10 ranking molecules with 90% accuracy (9 out of 10 molecules) as shown in Table 4.2.

Table 4.2: Verification of the Proposed Methodology with a small fragment library.

Sorting [a]	ID	Calculated (AD4) Total Energy	R1	R2	R3	R4	Estimated (GAMS) Total Energy
3	M_135	-25,14	1	5	15	5	-25.512
1	M_171	-26,81	22	5	15	5	-25.353
6	M_138	-24,46	1	5	17	5	-24.997
13	M_078	-23,97	1	5	4	5	-24.895
4	M_150	-25,00	22	5	17	5	-24.838
9	M_183	-24,28	17	5	15	5	-24.801
5	M_087	-24,78	22	5	4	5	-24.736
7	M_075	-24,35	1	5	1	5	-24.653
2	M_084	-25,25	22	5	1	5	-24.494
8	M_145	-24,30	30	5	15	5	-24.377

[a] sorting with respect to calculated total energies among all (150) of the molecules in the verification library.

On the Table 4.2, fragments used to generate corresponding small molecules were shown at the columns 4-7. Respective fragments can be seen in Table3.1. Detailed information about binding and docking energies of top ranking molecules in the verification library was shown in Table 4.3. Tools used in the calculation of binding and docking energies were indicated in parenthesis.

Table 4.3: Calculated (by AD4) and Estimated (by GAMS) values for the binding and docking energies of the top ranking molecules in the verification library.

Sorting [a]	ID	BE (AD4)	DE (AD4)	BE (GAMS)	DE (GAMS)
3	M_135	-11,16	-13,98	-11.383	-14.129
1	M_171	-11,99	-14,82	-11.246	-14.107
6	M_138	-11,48	-12,98	-11.737	-13.259
13	M_078	-10,53	-13,44	-11.022	-13.873
4	M_150	-11,66	-13,34	-11.599	-13.238
9	M_183	-10,73	-13,55	-11.002	-13.799
5	M_087	-10,83	-13,95	-10.885	-13.851
7	M_075	-10,88	-13,47	-11.055	-13.598
2	M_084	-11,22	-14,03	-10.918	-13.576
8	M_145	-10,74	-13,56	-10.801	-13.576

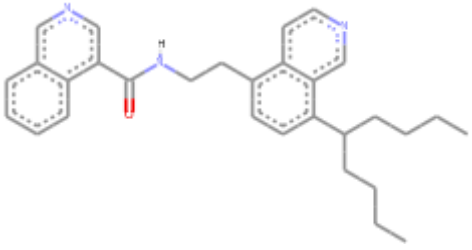
[a] sorting with respect to calculated total energies among all (150) of the molecules in the verification library.

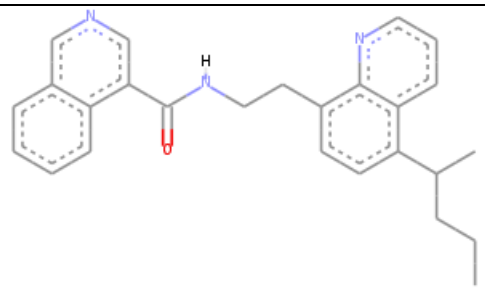
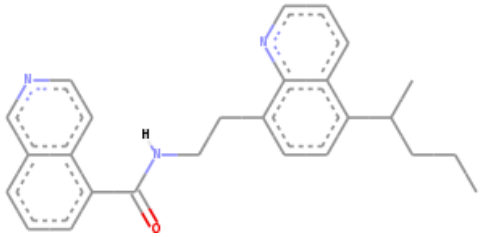
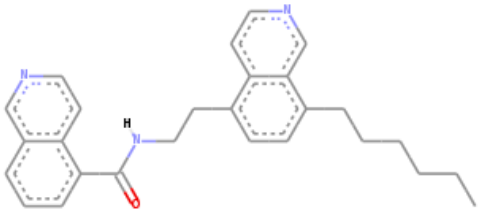
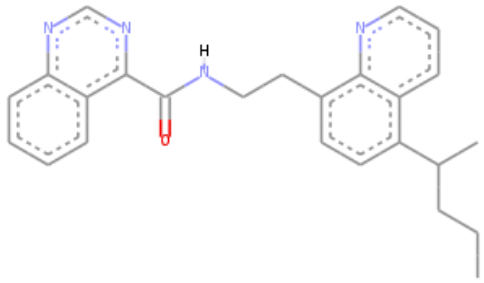
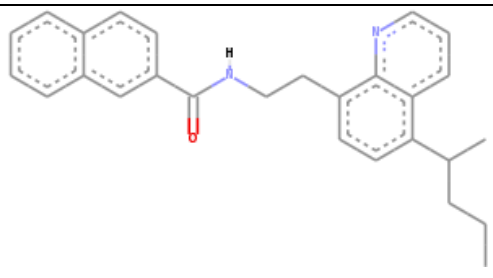
Although sorting of top ranking molecules differ between AD4 and GAMS derived energies, GAMS energy estimations show small deviations from energies calculated by AD4. Since 9 of the top 10 molecules was identified with small deviations in the binding and docking energies, the proposed model is applied to a large library of fragments in order to identify top scoring molecules that can be possible generated by the predetermined fragments.

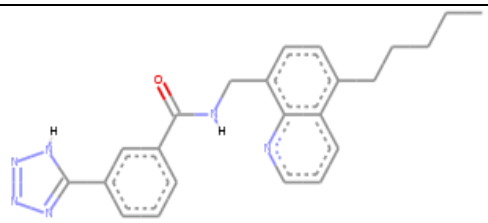
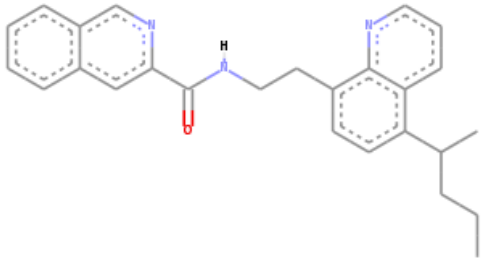
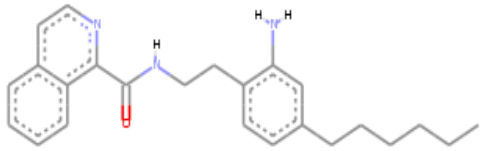
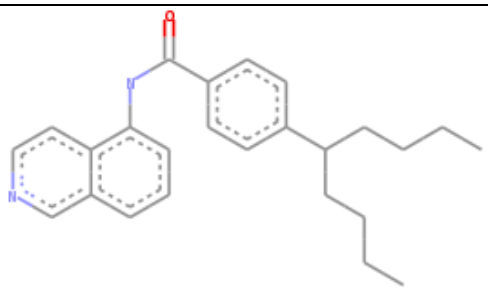
4.6 the Test Set of the Model

621 molecules having 4 different fragment sites with randomly chosen fragments were designed. Fragment library used in this process was shown in Table 3.1. Among ~600 designed molecules in the test library, top 10 molecules that were originally designed can be seen in Table 4.4. Contributions (weight coefficients) of each fragment in the 4 fragment sites to the total binding and docking energies were calculated with the proposed model. With respect to weight coefficients calculated, the top ranking molecules were estimated and drawn by Discovery Studio.

Table 4.4: Top 10 molecules of the original dataset are listed. Sorting is done according to total binding and docking energies. BE: Binding Energy, DE: Docking Energy

ID	**BE (AD4)**	**DE (AD4)**	**R1**	**R2**	**R3**	**R4**	**Structure**
dock06	-11,33	-15,56	27	5	4	3	

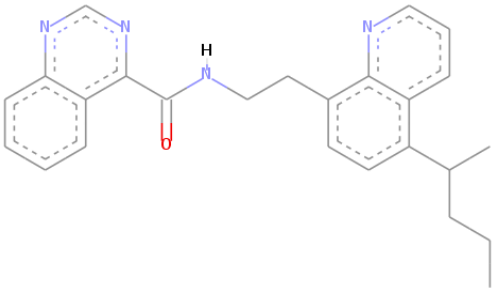
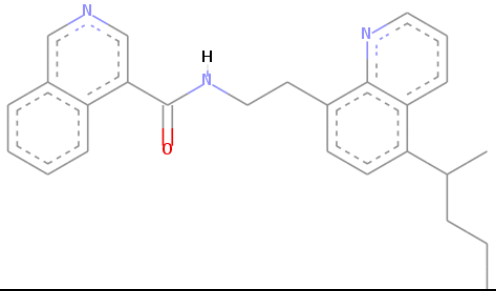
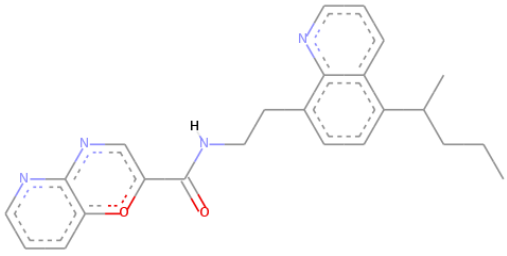
cluster045	-11,93	-14,65	27	5	6	8	
cluster047	-11,87	-14,62	6	5	6	8	
bind06	-11,53	-14,82	6	5	4	27	
cluster041	-11,59	-14,48	12	5	6	8	
cluster042	-11,6	-14,38	7	5	6	8	

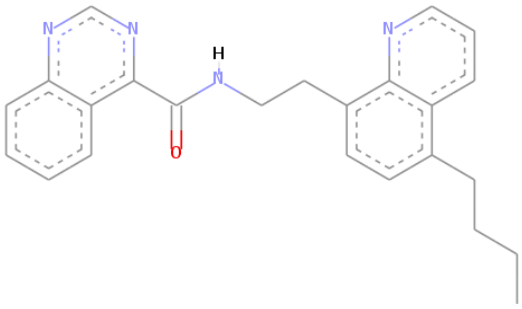
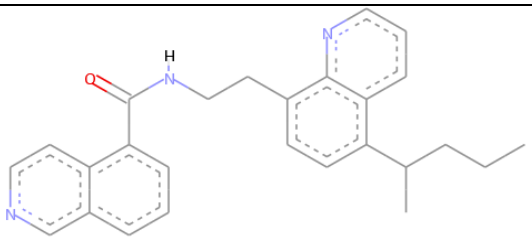
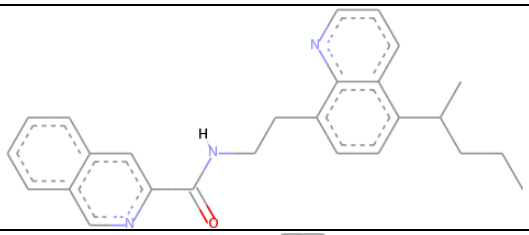
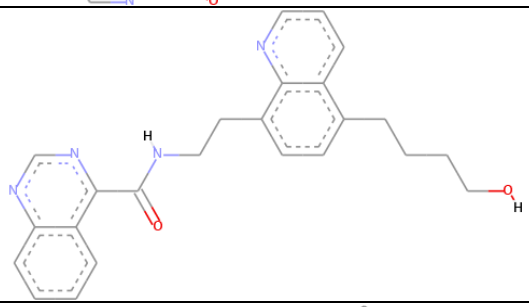
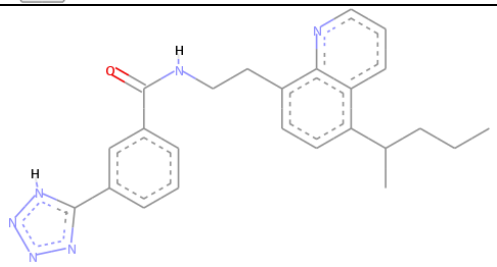
cluster011	-11,38	-14,5	11	4	6	27	
cluster044	-11,5	-14,34	9	5	6	8	
total04	-10,97	-14,58	22	5	2	27	
EA23	-11,17	-14,36	6	2	1	3	

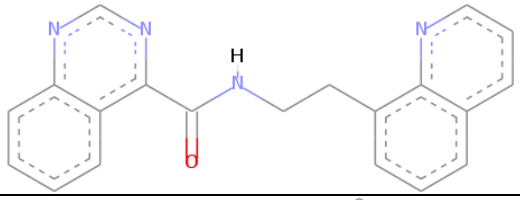
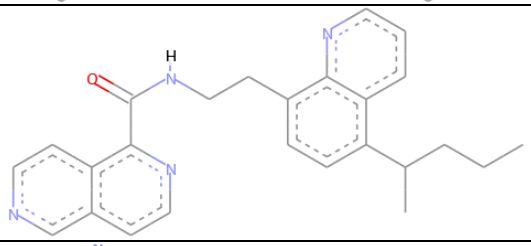
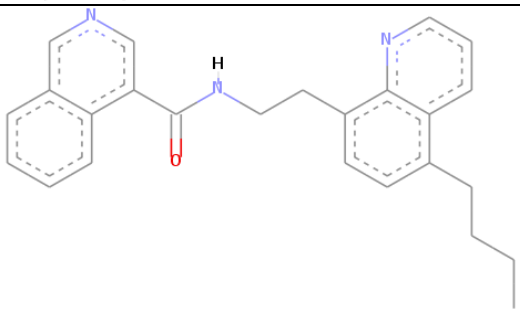
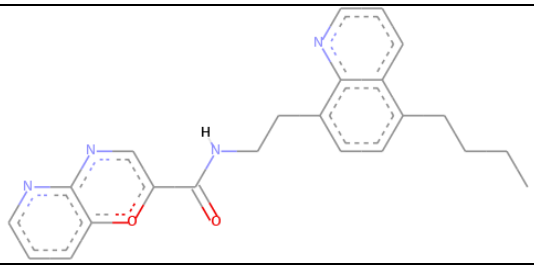
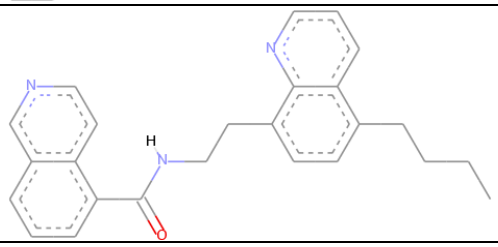
In the test set, ~600 combinations among over 80,000 possible combinations of fragments were generated randomly. This dataset was used in identification of better, even best, molecules that can be generated by combinations of given fragments (Table 3.1). Though neighboring fragments sites might possess structures containing aromatic rings, total number of rings in an aromatic group is restricted to two since none of the original molecules contain aromatic groups with three rings and also to limit molecular weights of the compounds. Estimation of top ranking molecules was achieved by python scripting as

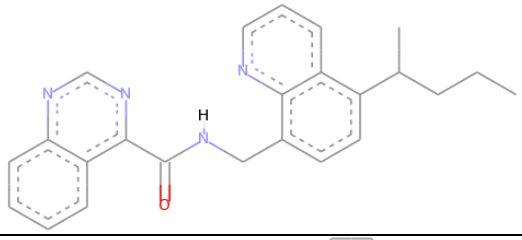
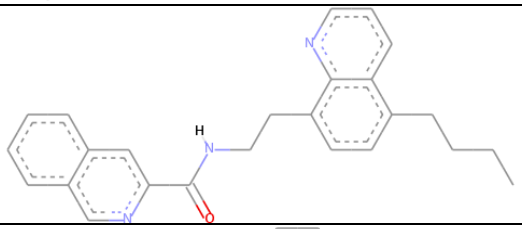
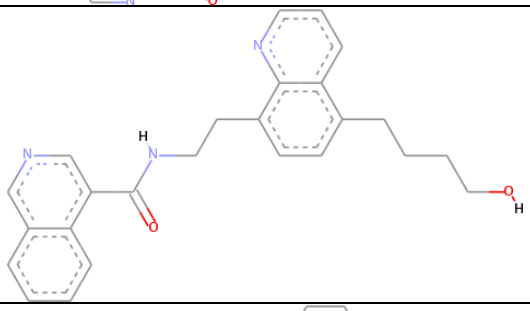
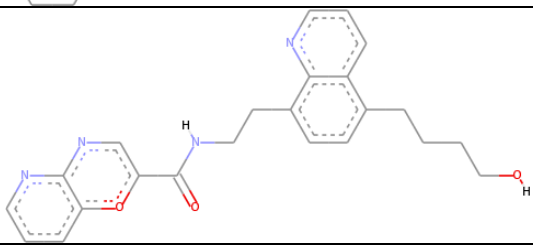
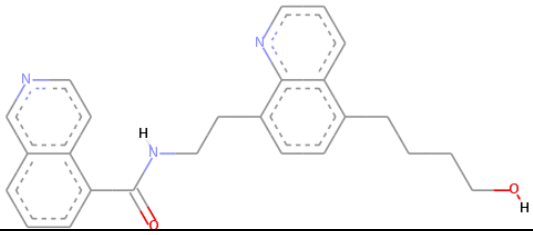
the summation of weight coefficients in the corresponding fragment sites with a restriction to total number of rings. A list of top 20 compounds with the estimation of lowest total binding and docking energies is listed on Table 4.5.

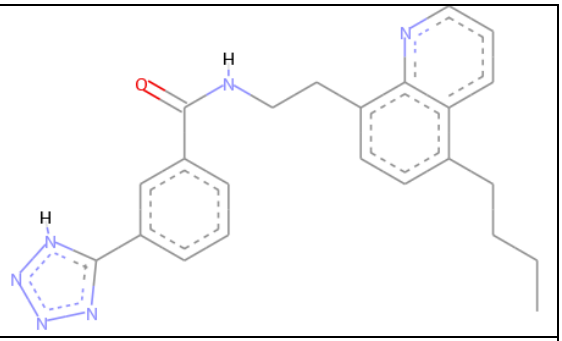
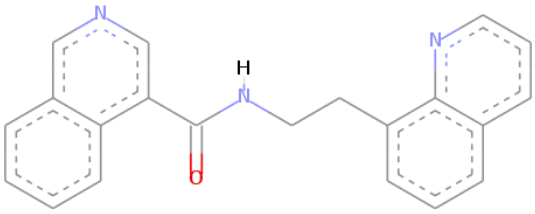
Table 4.5: List of top 20 molecules identified by the Model using GAMS..

ID	BE (AD4)	DE (AD4)	BE (GAMS)	DE (GAMS)	Structure
1	-11,59	-14,48	-11,898	-15,486	
2	-11,93	-14,65	-11,935	-15,962	
3	-11,6	-14,38	-11,855	-15,019	

4	-11,27	-14,03	-11,949	-14,914	
5	-11,87	-14,62	-11,845	-14,976	
6	-11,77	-14,6	-11,761	-14,939	
7	-9,93	-13,27	-12,432	-14,219	
8	-12,02	-15,26	-11,468	-15,14	

9	-10,00	-11,64	-11,538	-15,019	
10	-11,33	-14,25	-11,580	-14,884	
11	-11,56	-14,14	-11,986	-14,39	
12	-10,20	-12,83	-11,906	-14,447	
13	-11,44	-14,09	-11,896	-14,404	

14	-10,25	-12,77	-11,451	-14,743	
15	-10,91	-13,62	-11,812	-14,367	
16	-11,29	-14,44	-12,469	-13,695	
17	-10,78	-14,01	-12,389	-13,752	
18	-10,36	-13,57	-12,379	-13,709	

19	-10,56	-14,65	-11,519	-14,568	
20	-10,56	-12,02	-11,575	-14,495	

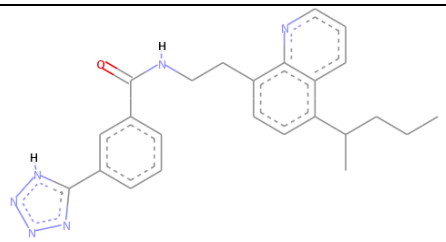
The algorithm could identify a molecule that is already included in the initial dataset since it is not restricted to do so. The model suggested 20 novel molecules, some of which are not included in the initial dataset, as having the highest total binding and docking energies. 6 of the top 10 molecules of the initial dataset were included in the top 20 estimations of the model designed. This finding also indicates the efficiency of the model in estimation of top ranking novel molecules from a given set of molecules. Both estimations and calculations of binding and docking energies of newly identified molecules were shown in Columns 2-5 on Table 4.5.

Although in the verification set, deviations of estimated binding and docking energies from calculated ones were small (max ~6%), it was greater in the test set as even ~25% for some specific cases, i.e. molecules 7, 9 and 14 in Table 4.5. Those molecules have a common property which is the 12th fragment in their R1 position. The only miscalculation of weight coefficients was seen in this specific case that caused an error rate of at max ~25% which is highly significant. This error rate might be due to relatively small number of molecules in the test dataset which is 600 over 80,000 possible combinations

whereas it was 60 over 150 possible combinations in the verification set. Besides, other than 12th fragment in the R1 position, all the energy estimations of GAMS were correlated with the energies calculated by AD4. Hence, the algorithm is advantageous since calculations of weigh coefficients of other fragments were quite successful and suggested fragment combinations were involved in the formation of better ranking molecules than in the original dataset.

As shown in Table 4.1, binding energy of Abiraterone is around -12 kcal/mol and its docking energy is around -13 kcal/mol. Those values were set as the first threshold for these newly identified molecules in order to discover more specific and stronger inhibitors of CYP17. Only one (8th) of the newly identified molecules passed the first threshold. A summary of properties for molecule 8th can be found in Table 4.6. Binding conformation of molecule 8th can be found in the Figure 4.5. miLogP values and molecular weight of compounds were calculated by Molinspiration Cheminformatics online software.

Table 4.6: Molecule 8th with energy values above the first threshold

ID	**BE (AD4) / DE (AD4)**	**BE - Vina**	**Structure**	**miLogP**	**Molcular Weight**
8	-12,02 kcal/mol -15,26 kcal/mol	-8,7 kcal/mol		4.49	414.51

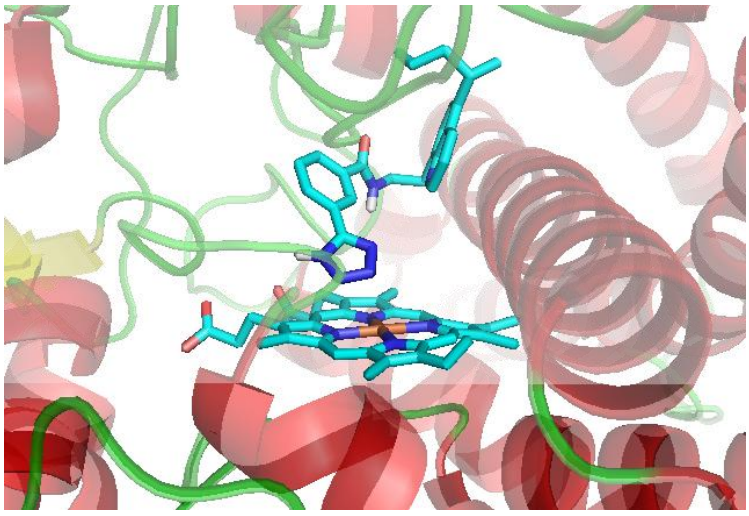
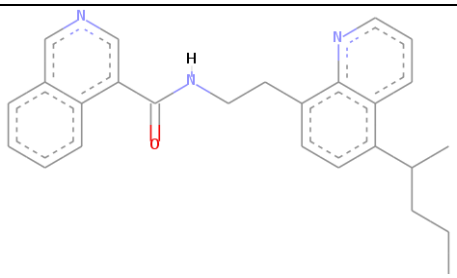
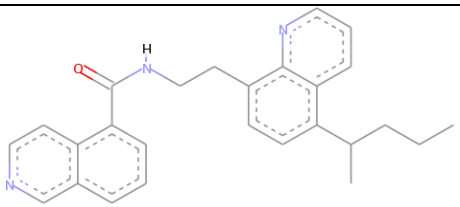
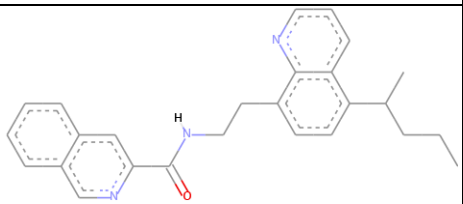
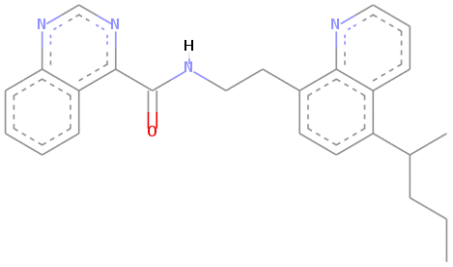
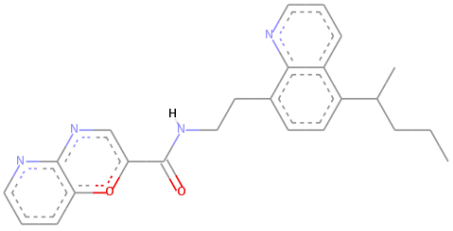


Figure 4.5: Binding conformation of the Molecule 8th estimated by ADT4

Since number of molecules discovered were limited to one with the first threshold, a second threshold with a greater binding energy were decided as binding energy smaller or equal to -11 kcal/mol and docking energy smaller or equal to -13 kcal/mol. Since binding energy of the pregnenolone and progesterone were around -10 - -11 kcal/mol, the second threshold also is legitimate so as to compete with natural substrates of CYP17. A summary of properties for the molecules with energy values above the second threshold can be found in Table 4.7. Binding conformations of the molecules passing the second threshold can be found in the Figure 4.6.

Table 4.7: Molecules with energy values above the second threshold

ID	BE (AD4) / DE(AD4)	BE - Vina	Structure	miLogP	Molcular Weight
2	-11,93 -14,65	-9,3		5.18	397.52
5	-11,87 -14,62	-9,4		5.15	397.52
6	-11,77 -14,6	-9,2		5.66	397.52
1	-11,59 -14,48	-9		4.93	398.51
3	-11,6 -14,38	-9,3		3.90	402.50

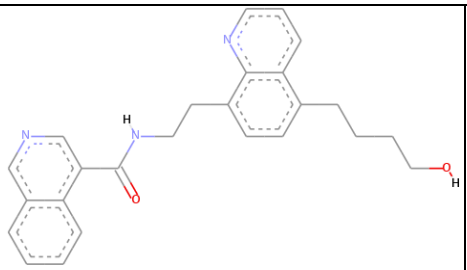
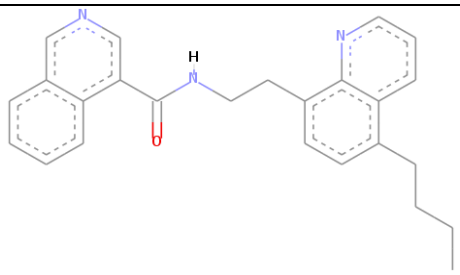
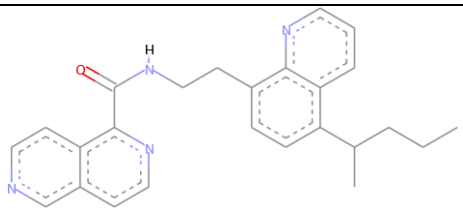
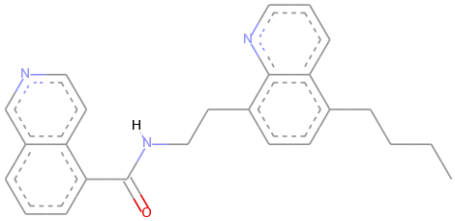
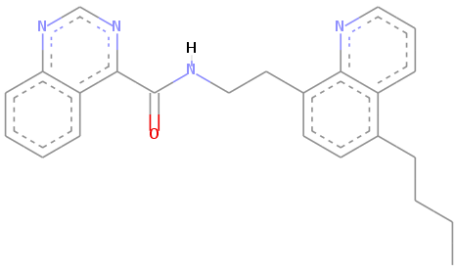
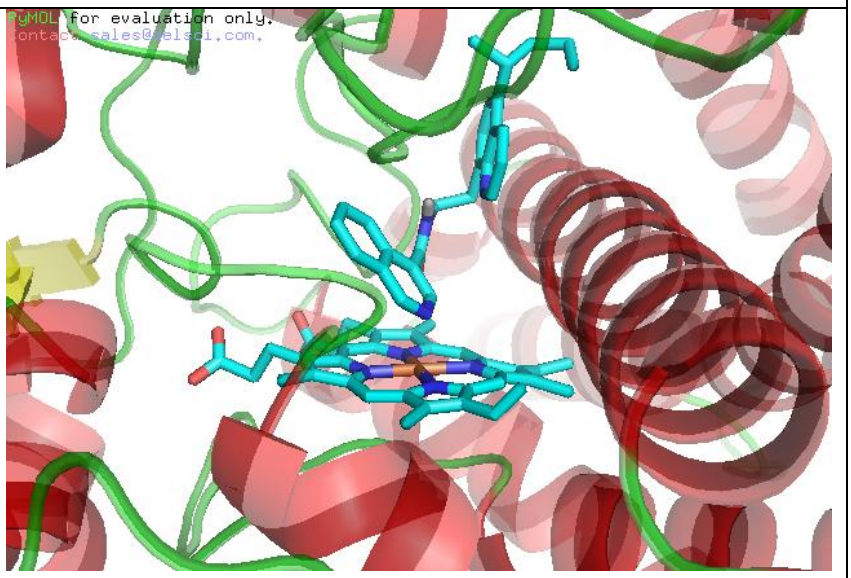
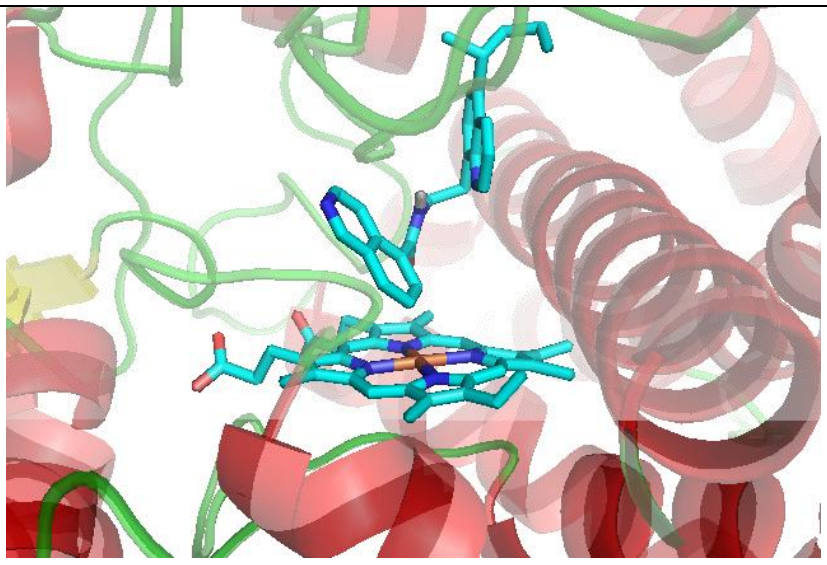
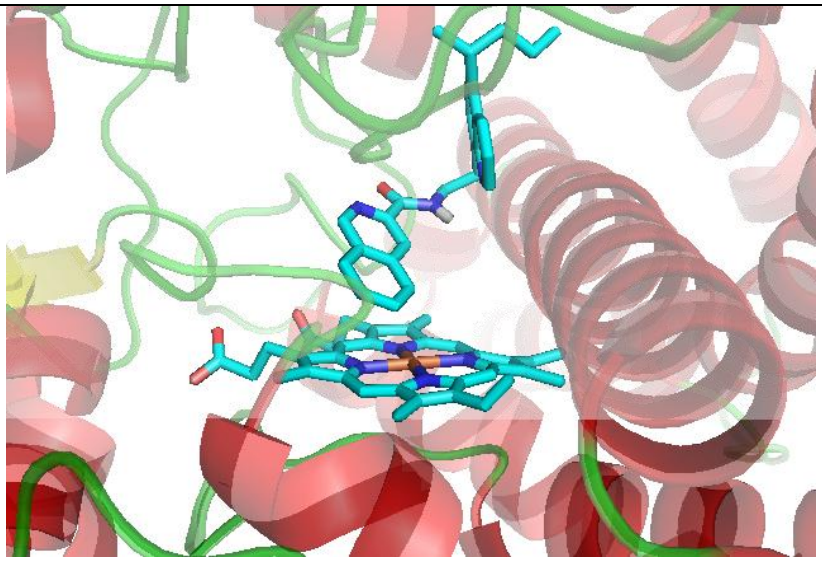
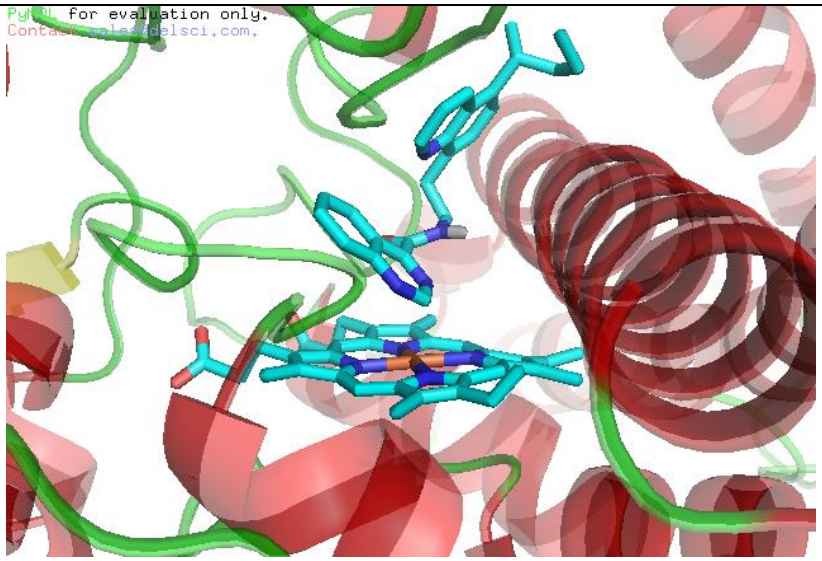
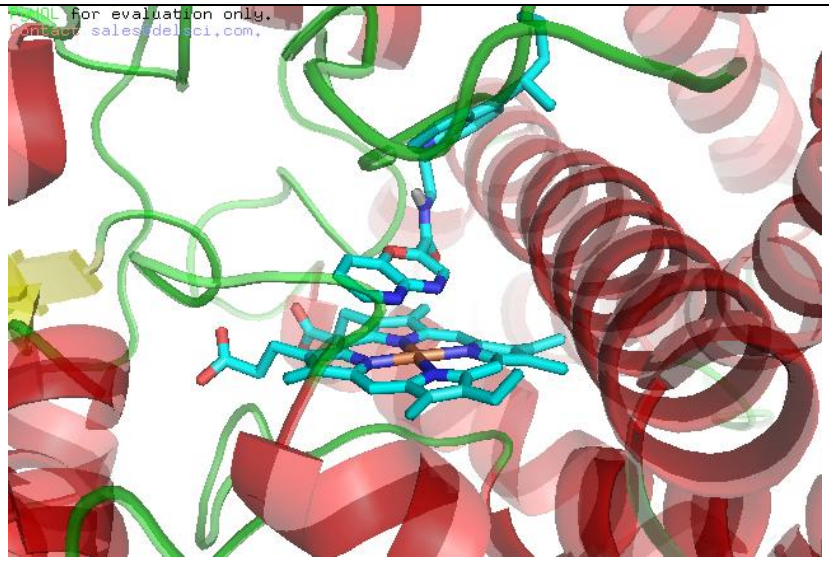
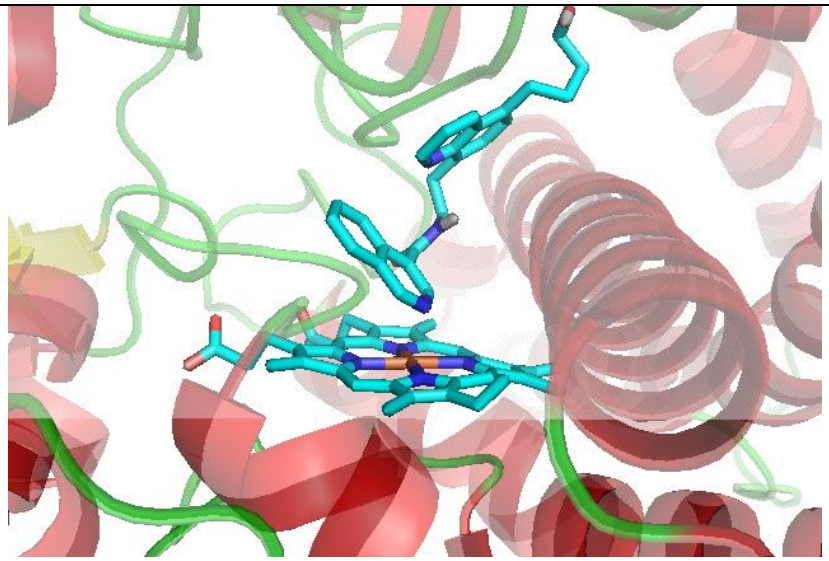
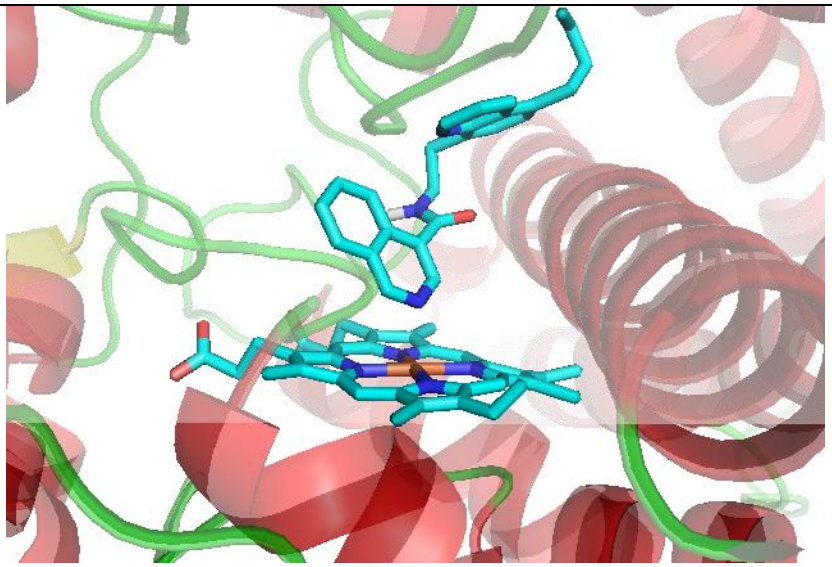
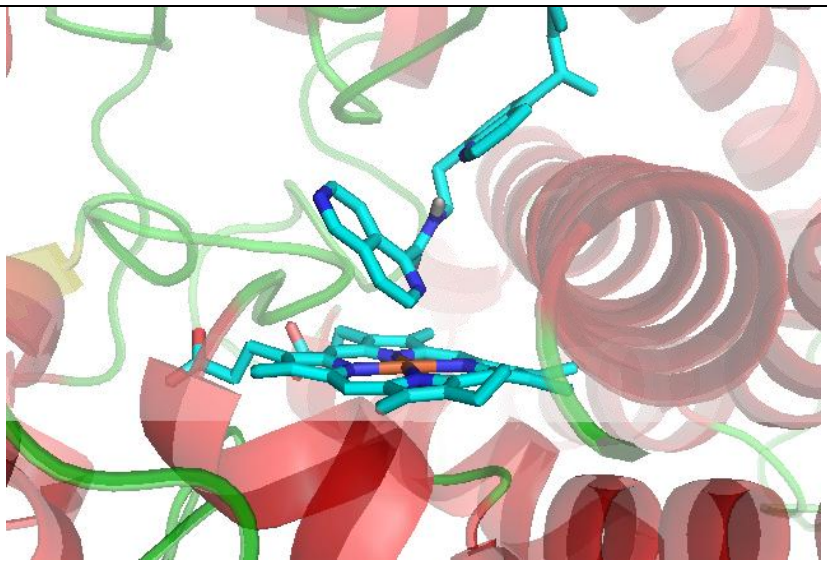
16	-11,29 -14,44	-10		3.58	399.49
11	-11,56 -14,14	-10		4.54	369.47
10	-11,33 -14,25	-9,2		4.34	398.51
13	-11,44 -14,09	-10		5.10	383.50
4	-11,27 -14,03	-10,4		4.29	370.46

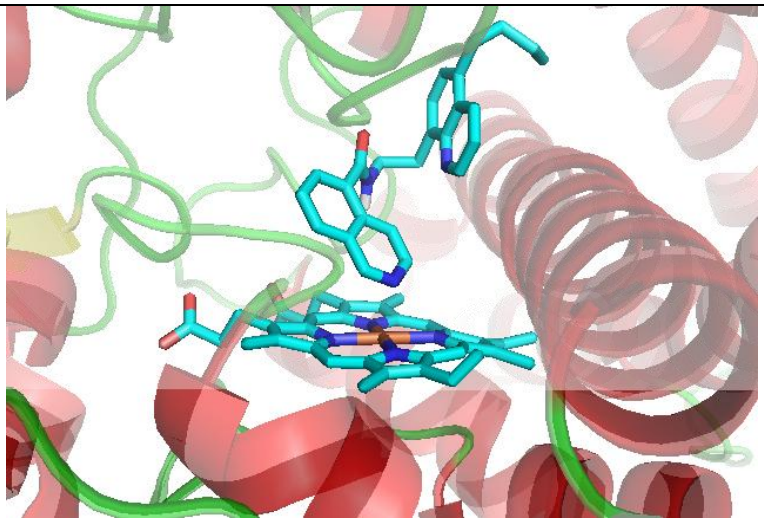
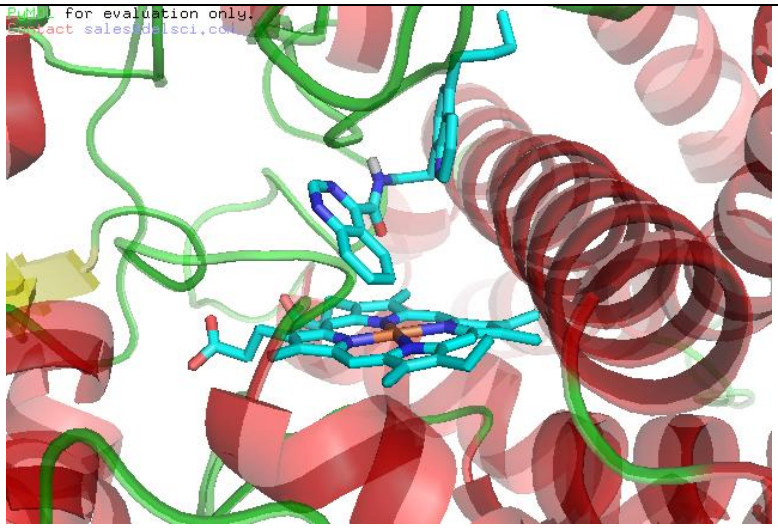
Table 4.8: Binding conformation of the newly discovered molecules with energy values above the second threshold, sorted by total binding and docking energies calculated by AD4.

ID	Structure
2	
5	

6	
1	

3	
16	

11	
10	

13	
4	

To compare affinities calculated by autodock4, newly discovered molecules and known drugs were screened with an alternative software called Vina. Energies calculated by Vina can be found in the third column on the table 4.6 and table 4.7. Energies of natural substrates and known drugs were also calculated by Vina as shown in Table 4.9. Vina screening also demonstrates that energy values of all newly discovered molecules are lower than ketoconazole and are compatible with the natural substrates.

Table 4.9: Binding energies of known inhibitors and natural substrates of CYP17 calculated by Vina

Inhibitors	**BE (Vina)**	**Substrates**	**BE (Vina)**
Ketoconazole	-7	Pregnenolone	-10,5
Abiraterone	-11,1	Progesterone	-10,6

Molecules having binding energy lower than -11 kcal/mol were selected and listed in the table 4.6 and table 4.7. All molecules bond to CYP17 in a position that is close to the iron atom of the heme group, which is a desired property for ideal inhibitors. With this fragment-based approach, in total, 11 novel molecules against CYP17 having smaller binding and docking energies than its natural substrates were discovered. By their low binding and docking energies these molecules were also comparable molecules to Abiraterone, the best drug known so far targeting CYP17.

Chapter 5:

Conclusions

Computational tools were used in this thesis in order to develop new therapeutics against prostate cancer using a new fragment-based drug design approach. The idea behind this proposed approach is to build novel compounds based on an initial scaffold and to improve affinity of drug candidates with the identification of the best combinations of fragments.

The treatment strategy for PC was decided as elimination of serum androgen levels by targeting androgen biosynthesis pathway. CYP17 protein was chosen as the protein target due to its importance in androgen biosynthesis. Structure of the enzyme was obtained from protein data bank and molecular dynamics simulation was performed to improve and validate its structure for later docking studies. Once RMSD of the protein converged to a stable value, indicating a stable structure, docking studies were performed.

Docking studies initially validated by docking natural substrates and known drugs to CYP17. The initial docking studies showed accuracy of the prediction of binding conformations of substrates into protein. Also binding and docking energies indicating strong affinity towards CYP17 was decided with this initial docking. Provided energy values were used in determination of threshold values for binding and docking energies of the newly discovered molecules and were used to compare their affinity and specificity for CYP17.

The fragment based de novo drug design model were generated, which uses a common scaffold with predetermined fragment sites on it based on a lead compound

discovered in another study. Fragments were determined by chemicals used against CYP17 in earlier studies or in studies of other groups. Once a fragment library was built, different combinations of fragments were combined randomly to develop new drugs based on a common scaffold with different fragment combinations.

The model composed of two stages, first stage is the training stage where contributions (weight coefficients) of each fragments calculated and the second stage is the test stage where new molecules were generated with the combinations of fragments having the best weight coefficients. In the first stage, the model requires two specifications: energy values of molecules and existence parameters of fragments in the corresponding molecule in order to estimate contribution of different variables into energies. In the second stage, estimated weigh coefficients was used to combine best fragments in order to develop the optimum structure possible with the given fragments. Before application of this model in a large fragment library, with which more than 80,000 molecules can be generated, a small portion of the initial fragment library was chosen in order to verify this fragment-based approach.

A verification library of 150 molecules was generated with a small number of fragments (varying from 2 to 5). All possible combinations of fragments included in this verification library. 60 of total 150 molecules were chosen randomly and the model tested on its accuracy. Since all molecules, possible with these fragments, were known, accuracy of the model in predicting top 10 molecules, with given combinations and respective energies, was determined. The model was successfully estimate binding and docking energies of top 10 molecules with small estimation errors and predicted the best 9 of top 10 molecules.

In the main part of this study, the optimum structure for CYP17 inhibitors to be composed of fragments in a given fragment library was decided without having all possible combinations of fragments. At the training stage, ~600 molecules (of over 80,000

possibilities) were generated and weight coefficients of each fragment were estimated. At the test stage, top 20 molecules were generated based on weight coefficients calculated. Estimations of the energy values based on calculated weight coefficients showed significant deviations from calculations of energies by AutoDock. Although the model failed in prediction of exact values for binding and docking energies, new molecules constructed by the combinations of predicted fragments were involved in the formation of better ranking molecules from the ones in the original dataset.

In the verification step 60 over 150 molecules gave promising results (at max ~6% deviations) whereas in the main step 600 over 80,000 molecules resulted in decreased accuracy in energy estimations (at max ~25% deviations for a specific fragment). The main reason for the deviations in the energy estimations is the limited number of molecules in the initial dataset. Better estimations were expected with increasing number of inputs. The number of inputs can either be increased by adding more molecules into initial dataset or by adding more constraints like IC50 values or calculations from other computational tools.

In this study, fragment-based drug design approach was used in order to develop novel compounds against Cyp17. A non-linear programming problem was described and solved using GAMS in order to identify best combination of fragments. A grid-based docking strategy was used for identification and analysis of generated compound. The results revealed eleven novel molecules with better binding and docking energies than natural substrates of CYP17 and also comparable with Abiraterone, a known drug for CYP17.

In the future studies, more extensive fragment libraries and larger dataset of molecules can be used with the same scaffold and same procedure in order to increase probability of identifying better molecules with an increasing accuracy. The scaffold can be modified to reduce fragment sites since most of the top molecules have the same linker

region. Since energy values might differ with different computational tools due to differences in search algorithm and/or scoring functions, other computational tools can be used too in order to discover novel molecules with the same procedure.

Supplementary Materials

Supplementary 1: Configuration file of the Minimization for NAMD simulation

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization of CYP17 in Water Box
# Minimization (w/ fixed atoms), heating at NVT, production
# run at NPT

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          ubqn_wb.psf
coordinates        ubqn_wb.pdb

set temperature    0
temperature        $temperature
outputname         out1
restartname        res1
#firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input

paraTypeCharmm          on
parameters          par_all27_prot_lipid.inp
parameters          par_custom_heme.inp

# Force-Field Parameters

exclude             scaled1-4
1-4scaling          1.0
```

```
cutoff              12
switching           on
switchdist          8
pairlistdist        13.5

# Integrator Parameters

timestep            2
rigidbonds          all
rigidTolerance      0.00000001
nonbondedFreq       1
fullElectFrequency  2
stepsPerCycle       10

# Constant Temperature Control

langevin            on
langevinDamping     5
langevinTemp        310
langevinHydrogen    off

# Periodic Boundary Conditions

cellBasisVector1    84.0   0.0   0.0
cellBasisVector2     0.0  82.0   0.0
cellBasisVector3     0.0   0.0  85.0
cellOrigin          -1.36 -8.94 1.80

margin              2.5
wrapAll             on


# PME (for full-system periodic electrostatics)

PME                 on
PMEGridSpacing      1.0

#PMEGridSizeX        44
#PMEGridSizeY        82
```

```
#PMEGridSizeZ       85

# Constant Pressure Control (variable volume)

useGroupPressure      yes
useFlexibleCell             no
useConstantRatio      no

langevinPiston        on
langevinPistonTarget  1.01325
langevinPistonPeriod  100
langevinPistonDecay   50
langevinPistonTemp    310

# Output

outputEnergies      500
outputPressure      500
xstFreq             500
DCDFreq             500
restartfreq         500

#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################

# Fixed Atoms

fixedAtoms          on
fixedAtomsForces    on
fixedAtomsFile      fix_all.pdb
fixedAtomsCol       B


#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

# run one step to get into scripting mode
```

```
minimize            0

# turn off until later

langevinPiston      off

# minimize nonbackbone atoms

minimize            10000
output              min_fix

# min all atoms

fixedAtoms          off
minimize            10000
output              min_all

# Increase the temperature gradually

for { set TEMP 10 } { $TEMP < 310 } { incr TEMP 10 } {
    langevinTemp  $TEMP
    run           5000
}

langevinTemp        310
output              heating

langevinPiston      on

run 10000
```

Supplementary 2: Configuration file of the Production run for NAMD simulation

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Equilibration of CYP17 in Water Box
# production  run at NPT

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure         ubqn_wb.psf
coordinates       ubqn_wb.pdb
binCoordinates    out1.coor
binVelocities     out1.vel
extendedSystem    out1.xsc

set temperature   310
outputname        out2
restartname       run2
firsttimestep     280000
temperature       $temperature

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input

paraTypeCharmm          on
parameters          par_all27_prot_lipid.inp
parameters          par_custom_heme.inp

# Force-Field Parameters

exclude             scaled1-4
1-4scaling          1.0
```

```
cutoff              12
switching           on
switchdist          8
pairlistdist        13.5

# Integrator Parameters

timestep            2
rigidbonds          all
rigidTolerance      0.00000001
nonbondedFreq       1
fullElectFrequency  2
stepsPerCycle       10

# Constant Temperature Control

langevin            on
langevinDamping     5
langevinTemp        310
langevinHydrogen    off

# Periodic Boundary Conditions

cellBasisVector1    84.0   0.0   0.0
cellBasisVector2     0.0  82.0   0.0
cellBasisVector3     0.0   0.0  85.0
cellOrigin          -1.36 -8.94 1.80

margin              2.5
wrapAll             on

# PME (for full-system periodic electrostatics)

PME                 on
PMEGridSpacing      1.0
#PMEGridSizeX        44
#PMEGridSizeY        82
#PMEGridSizeZ        85
```

```
# Constant Pressure Control (variable volume)

useGroupPressure      yes
useFlexibleCell             no
useConstantRatio      no

langevinPiston        on
langevinPistonTarget  1.01325
langevinPistonPeriod  100
langevinPistonDecay   50
langevinPistonTemp    310

# Output

outputEnergies      1000
outputPressure      1000
xstFreq           5000
DCDFreq             5000
restartfreq         5000

#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################

# Fixed Atoms

#fixedAtoms          on
#fixedAtomsForces    on
#fixedAtomsFile      fix_all.pdb
#fixedAtomsCol       B

#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################


run 5000000
```

Bibliography

1. Klinke, D.J., 2nd, Enhancing the discovery and development of immunotherapies for cancer using quantitative and systems pharmacology: Interleukin-12 as a case study. J Immunother Cancer, 2015. 3: p. 27.
2. Andreoli, F. and A. Del Rio, Computer-aided Molecular Design of Compounds Targeting Histone Modifying Enzymes. Comput Struct Biotechnol J, 2015. 13: p. 358-65.
3. Clark, D.E. and S.D. Pickett, Computational methods for the prediction of 'drug-likeness'. Drug Discov Today, 2000. 5(2): p. 49-58.
4. Anderson, A.C., The process of structure-based drug design. Chem Biol, 2003. 10(9): p. 787-97.
5. Organization, W.H. 2015: http://www.who.int/topics/cancer/en/.
6. Health, D.o.C.T.M.o., Turkish Cancer Statistics Database. 2009.
7. Armutlu, P., Ozdemir, M., Ozdas, S., Kavakli, I., Turkay, M., Discovery of novel cyp17 inhibitors for the treatment of prostate cancer with structure-based drug design. Letters in Drug Design, 2009. 38(Discovery 6(5)): p. 337-344.
8. Acar, O., T. Esen, and N.A. Lack, New therapeutics to treat castrate-resistant prostate cancer. ScientificWorldJournal, 2013. 2013: p. 379641.
9. Karlou, M., V. Tzelepi, and E. Efstathiou, Therapeutic targeting of the prostate cancer microenvironment. Nat Rev Urol, 2010. 7(9): p. 494-509.
10. Clement, O.O., et al., Three dimensional pharmacophore modeling of human CYP17 inhibitors. Potential agents for prostate cancer therapy. J Med Chem, 2003. 46(12): p. 2345-51.
11. Handratta, V.D., et al., Novel C-17-heteroaryl steroidal CYP17 inhibitors/antiandrogens: synthesis, in vitro biological activity, pharmacokinetics,

and antitumor activity in the LAPC4 human prostate cancer xenograft model. J Med Chem, 2005. 48(8): p. 2972-84.

12. Vasaitis, T.S., R.D. Bruno, and V.C. Njar, CYP17 inhibitors for prostate cancer therapy. J Steroid Biochem Mol Biol, 2011. 125(1-2): p. 23-31.
13. Yap, T.A., et al., Targeting CYP17: established and novel approaches in prostate cancer. Curr Opin Pharmacol, 2008. 8(4): p. 449-57.
14. Singh, A.S., et al., Mechanisms of disease: Polymorphisms of androgen regulatory genes in the development of prostate cancer. Nat Clin Pract Urol, 2005. 2(2): p. 101-7.
15. Labrie, F., et al., Comparable amounts of sex steroids are made outside the gonads in men and women: strong lesson for hormone therapy of prostate and breast cancer. J Steroid Biochem Mol Biol, 2009. 113(1-2): p. 52-6.
16. Holzbeierlein, J., et al., Gene expression analysis of human prostate carcinoma during hormonal therapy identifies androgen-responsive genes and mechanisms of therapy resistance. Am J Pathol, 2004. 164(1): p. 217-27.
17. Jernberg, E., et al., Characterization of prostate cancer bone metastases according to expression levels of steroidogenic enzymes and androgen receptor splice variants. PLoS One, 2013. 8(11): p. e77407.
18. Bruno, R.D. and V.C. Njar, Targeting cytochrome P450 enzymes: a new approach in anti-cancer drug development. Bioorg Med Chem, 2007. 15(15): p. 5047-60.
19. Freire, E., A thermodynamic approach to the affinity optimization of drug candidates. Chem Biol Drug Des, 2009. 74(5): p. 468-72.
20. Stein, M.N., et al., Androgen synthesis inhibitors in the treatment of castration-resistant prostate cancer. Asian J Androl, 2014. 16(3): p. 387-400.
21. Lipinski, C.A., Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods, 2000. 44(1): p. 235-49.

22. Alonso, H., A.A. Bliznyuk, and J.E. Gready, Combining docking and molecular dynamic simulations in drug design. Med Res Rev, 2006. 26(5): p. 531-68.
23. Murray, C.W. and D.C. Rees, The rise of fragment-based drug discovery. Nat Chem, 2009. 1(3): p. 187-92.
24. Congreve, M., et al., A 'rule of three' for fragment-based lead discovery? Drug Discov Today, 2003. 8(19): p. 876-7.
25. Chen, X., et al., A ligand-observed mass spectrometry approach integrated into the fragment based lead discovery pipeline. Sci Rep, 2015. 5: p. 8361.
26. Mashalidis, E.H., et al., A three-stage biophysical screening cascade for fragment-based drug discovery. Nat Protoc, 2013. 8(11): p. 2309-24.
27. La, J., et al., Identification of mechanistically distinct inhibitors of HIV-1 reverse transcriptase through fragment screening. Proc Natl Acad Sci U S A, 2015. 112(22): p. 6979-84.
28. Schneider, G. and H.J. Bohm, Virtual screening and fast automated docking methods. Drug Discov Today, 2002. 7(1): p. 64-70.
29. Rashid, M.M. and D.F. Hayes, Nonlinear programming technique for analyzing flocculent settling data. Water Environ Res, 2014. 86(4): p. 346-59.
30. Karplus, M. and J.A. McCammon, Molecular dynamics simulations of biomolecules. Nat Struct Biol, 2002. 9(9): p. 646-52.
31. Papaleo, E., Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity. Front Mol Biosci, 2015. 2: p. 28.
32. McCammon, J.A., B.R. Gelin, and M. Karplus, Dynamics of folded proteins. Nature, 1977. 267(5612): p. 585-90.
33. Case, D.A., et al., The Amber biomolecular simulation programs. J Comput Chem, 2005. 26(16): p. 1668-88.

34. Brooks, B.R., et al., CHARMM: the biomolecular simulation program. J Comput Chem, 2009. 30(10): p. 1545-614.
35. Christen, M., et al., The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem, 2005. 26(16): p. 1719-51.
36. Phillips, J.C., et al., Scalable molecular dynamics with NAMD. J Comput Chem, 2005. 26(16): p. 1781-802.
37. Knapp, B., et al., vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. J Comput Chem, 2010. 31(16): p. 2868-73.
38. Kitchen, D.B., et al., Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov, 2004. 3(11): p. 935-49.
39. MacKerell, A.D., et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B, 1998. 102(18): p. 3586-616.
40. Autenrieth, F., et al., Classical force field parameters for the heme prosthetic group of cytochrome c. J Comput Chem, 2004. 25(13): p. 1613-22.
41. Collins, J.R.C., D. L.; Loew, G. H., Valproic acid metabolism by cytochrome P450: a theoretical study of stereoelectronic modulators of product distribution. J. Am. Chem. Soc., 1991. 113(7): p. 2736-2743.
42. Biesiada, J., et al., Survey of public domain software for docking simulations and virtual screening. Hum Genomics, 2011. 5(5): p. 497-505.
43. Ashtawy, H.M. and N.R. Mahapatra, Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. BMC Bioinformatics, 2015. 16 Suppl 6: p. S3.
44. Wang, R., L. Lai, and S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des, 2002. 16(1): p. 11-26.

45. Huang, S.Y. and X. Zou, An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. J Comput Chem, 2006. 27(15): p. 1876-82.

46. Morris, G.M., et al., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem, 2009. 30(16): p. 2785-91.

47. DeVore, N.M. and E.E. Scott, Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. Nature, 2012. 482(7383): p. 116-9.