Structure-Based Drug Design for Prostate Cancer

by

Muhittin Emre Özdemir

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

Master of Science
in
Computational Sciences and Engineering

Koç University

January 2008

Koc University
Graduate School of Sciences and Engineering

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

Muhittin Emre Özdemir

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)

_____

İbrahim Halil Kavaklı, Ph. D. (Co-Advisor)

_____

Özlem Keskin, Ph. D.

_____

Alkan Kabakçıoğlu, Ph. D.

_____

Deniz Yüret, Ph. D.

Date: _____

to my family…

ABSTRACT

Structure-Based Drug Design is a powerful method for designing inhibitors with high specificity. This method can be used for diseases where a single biochemical or structural function can be targeted for treatment. Structure-Based Drug Design has been successful in developing drug for several diseases involving hypertension, influenza, cancer and AIDS.

In this thesis, the objective is to design novel inhibitors for the treatment of prostate cancer using structure-based drug design approach. Prostate Cancer is the most common malignancy among males. There are several treatment procedures for prostate cancer. However, many cases are left untreated due to severe side effects and resulting decrease in quality of life. Nonetheless, inhibition of a single enzyme, CYP17, offers a specific treatment possibility with very low risk of side effects. Molecular dynamics methods are used to validate the model structure of CYP17. Computational tools for molecular docking are employed for discovering and developing novel agents inhibiting CYP17. Experimental approaches are followed for testing biological activity and toxicity of these molecules.

ÖZET

Yapıya dayalı ilaç dizaynı enzim aktivitelerine karşı yüksek seçicilikte engelleyici elde etmek için önemli bir yöntemdir. Bu yöntem kullanılarak belirli biyokimyasal ya da yapısal işlevler hedeflenebilir. Yapıya dayalı ilaç dizaynı yaklaşımı ile hipertansiyon, grip, kanser ve AIDS'in de aralarında bulunduğu birçok hastalığa yönelik ilaçlar başarıyla geliştirilmiştir.

Bu tez çalışmasında, yapıya dayalı ilaç dizaynı yöntemi ile prostat kanserinin tedavisine yönelik yeni ilaç adaylarının geliştirilmesi amaçlanmıştır. Prostat kanseri erkekler arasında en yaygın kanser türüdür. Bu hastalığın tedavisine yönelik çeşitli yöntemler geliştirilmiştir. Fakat birçok hasta için, bu yöntemlerin yan etkileri ve neden oldukları rahatsızlıklar kullanımlarını kısıtlamaktadır. Buna rağmen, CYP17 enziminin biyokimyasal işlevlerinin engellenmesi seçici ve yan etki riski düşük alternatif sunmaktadır. Moleküler dinamik yöntemleri kullanılarak CYP17 yapısı incelenmiştir. Hesaplamalı biyoloji araçları kullanılarak CYP17 ile en uygun tahmini etkileşimleri kuracak yeni moleküller bulunmuştur. Deneysel yöntemler kullanılarak, ilaç adaylarının CYP17 üzerindeki etkisi ve hücre üzerindeki zararlı etkisi incelenmiştir.

ACKNOWLEDGEMENTS

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

Dr. Metin Türkay and Dr. İ. Halil Kavaklı, my advisors, for their hard work and guidance throughout this entire thesis process and for believing in my abilities. I have learned so much from them, and thank you so much for this great experience.

Dr. Özlem Keskin, for participating in my thesis committee. Thank you for supporting and guiding me throughout the past two years.

Dr. Alkan Kabakçıoğlu and Dr. Deniz Yuret, for participating in my thesis committee. Thank you for your insight and guidance.

Dr Şule Beyhan Özdaş, for assisting in putting my research study into action. Thank you for your teachings and friendship.

Aytuğ Tunçel and Mehmet Ali Yatbaz for being with me in all good and bad times. My flatmates, Onur Unutulmaz, Mehmet Ali Dundar, Mehmet Akgul and Ergun Biçici, for sharing past two years with me.

My dear friend Pelin and my research group friends, Natali, Besray, Seda, Hande, Uğur, Pınar, Fadime, and Ali for their help and encouragement. My fellow graduate students, for their friendship.

My parents Emel and Muammer; my brother Hakan, for their never-ending love and support in all my efforts, and for giving me the foundation to be who I am. To them I dedicate this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Before 20th century, drugs were discovered, by trial and error, from roots, barks or leaves of plants to relieve pain or fever. In present, novel drug discovery and development is a scientific challenge, where biology, chemistry and computational methods are merged.

Combinatorial chemistry offers vast numbers of new molecules which can be used as drug candidates. Moreover, this large set of compounds became manageable with the use of high-throughput screening (HTS) technology for pharmaceutical companies. Nonetheless, operation of these experimental systems requires have high cost, and specificity of the compounds on investigated system can not be assessed.

Advances in molecular biology and X-Ray crystallography enabled virtual visualization of protein structures. Proteins resulting in biochemical or structural components of diseases are valuable data for drug design and discovery studies. Compounds providing a number of specific interactions with the backbone atoms and side chain atoms at the active site of the protein can be designed using structure-based drug design techniques. Using this method, specific drugs can be designed to inhibit certain enzymes without high costs.

Prostate cancer is the most common malignancy and age-related cause of cancer deaths among male worldwide. Although main cause of the disease has not been found, family history and increased age are among important risk factors. High level of androgens in blood is observed in many prostate cancer patients. Chemotherapy, radiation therapy, hormonal therapy ad surgical procedures such as prostatectomy and orchiectomy are most common treatment practices. However, all these methods suffer

side effects. Therefore, development of non-hormonal treatment methods with low risk of side effects is needed.

Androgens are major growth factors for prostate cancer cell. Hormonal therapies and orchiectomy are means of reducing blood levels these hormones. Testosterone and dehydro-testosterone are the most effective androgens on prostate growth. Biosynthesis of these molecules requires a key enzyme synthesized only in adrenal and testis tissues, CYP17. Specific inhibition of this enzyme, alone, would exert control over androgen synthesis without high side effect risk. Therefore, CYP17 is a good target for structure-based drug design.

CYP17 is a microsomal enzyme, catalyzing two key reactions in androgen biosynthesis pathway. Structure of this enzyme has not been resolved yet; however, a homology based model is generated. Based on this model, novel inhibitors for CYP17 can be designed and developed.

In this study, molecular dynamics simulations and docking tools are used to identify novel inhibitor compounds for CYP17 as potential treatment agents for prostate cancer.

Chapter 2 provides necessary background for prostate cancer and structure-based drug design. Computational and experimental methods used in this study are given in Chapter 3. Results and discussion are given in Chapter 5. This thesis is concluded with short summary of the work done, and recommendations on future research.

Chapter 2

Overview

2.1.   Prostate Cancer

2.1.1. Prostate and Prostate Cancer

The prostate is a male reproductive organ about 3 centimeters long and weighs about 20 grams in an adult. As a glandular tissue, the main function of this organ is to store and help produce seminal fluid. Prostate carcinomas are mostly adenocarcinomas, meaning glandular tissue is the origin of carcinoma. In the early phases, prostate cancer (PC) spreads to local nodes and structures. In the late phases, metastasis mostly occurs to local nodes, bone, supradiaphragmatic lymph nodes and lung. [1]

2.1.2. Epidemiology of Prostate Cancer

Prostate cancer is the most common malignancy among men in many countries (Figure 2.1). [2] Cancer statistics studies in US estimate 29% of all new male cancer diagnosis' to be PC. Although it is very improbable to develop PC before the age of 40, 1 out of 6 men will be diagnosed with invasive PC throughout their lifetime. Recently, PC has been ranked as the 2nd leading cause of cancer deaths among males.[3]

Incidence rate of PC holds a large variation among racial/ethnic groups. Among these groups African-American population has the highest incidence rate, followed by Caucasian; and the lowest rate is seen on Japanese population. These differentials among races are not completely understood but, dietary and culturally mediated differences as well as genetic factors have been suggested as possible explanations. [2-4] Activities of some androgen pathway enzymes have been found to differ between

African-American, Caucasian and Japanese subjects, which, can also be associated with the rate of PC. [5]
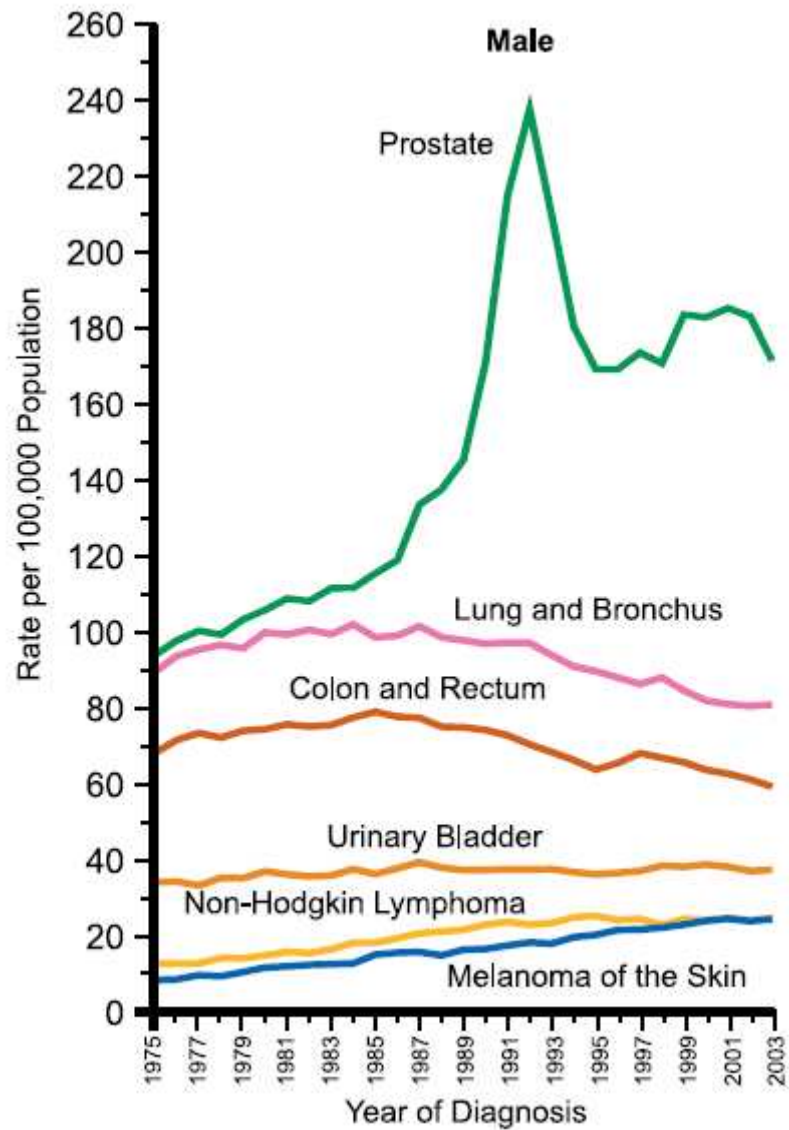
Figure 2.1: Annual Age-adjusted Male Cancer Incidence Rates, US [2]

2.1.3. Etiology of Prostate Cancer

The specific cause of prostate cancer remains unknown. Environment, diet, lifestyle, age and family history influence chance of developing PC; however, most important risk factors associated with PC are increased age and family history.[1] Although it is very unlikely to be diagnosed with PC before the age of 40, as men increase in age, their risk of developing PC increases exponentially. A study estimates 2 out of 5 men over the age of 50 to have slow growing PC that can be detected and this rate increases with age. [6] Having a family member with the history of PC increases an individual's risk factor by 2 fold, with two such relatives risk increases by 5 fold, and for three such relatives risk increases by 11 fold. [7]

2.1.4. Screening of Prostate Cancer

Prostate cancer accounts for more deaths for men than breast cancer does in women. Early diagnosis is very important for prostate cancer treatment.[1] When diagnosis is done in this stage treatments are very effective and population mortality can be reduced. However, most of the patients do not suffer symptoms in the early stages; and if diagnosis is done after symptoms are seen, metastasis is common. [8] Currently, two tests are common for first step screening of PC: PSA test and DRE. PSA, prostate specific antigen, is a glycoprotein mainly produced by prostate. Elevated levels of serum PSA is used as a sign of high PC risk. Although, PSA testing has improved early detection of PC, the method is not completely efficient. 20% of the PC patients have normal PSA levels, which decreases the efficiency of the test dramatically.[1, 8] Therefore, this test is combined with DRE, digital rectal examination. Patients with urological problems are subjected to these tests, and irregularities in prostate gland can be identified. Although 50% of irregularities are due to cancer, this combined method is the most effective screening tool up-to-date. [1]

2.1.5. Genetic and Metabolic background for Prostate Cancer

A positive family history is a strong risk factor for PC, where risk increases with the number of affected relatives increase. Therefore, hereditary components of the disease are studied widely. Several loci have been reported as human prostate cancer susceptibility regions, and several genes with their functions are studied. These studies help reveal components responsible for prostate cancer progression, as well as identify genetic markers for early detection. Figure 2.2 shows human prostate cancer susceptibility loci, where susceptibility genes are harbored. Since there are several loci related to PC, there is obvious heterogeneity. [9]
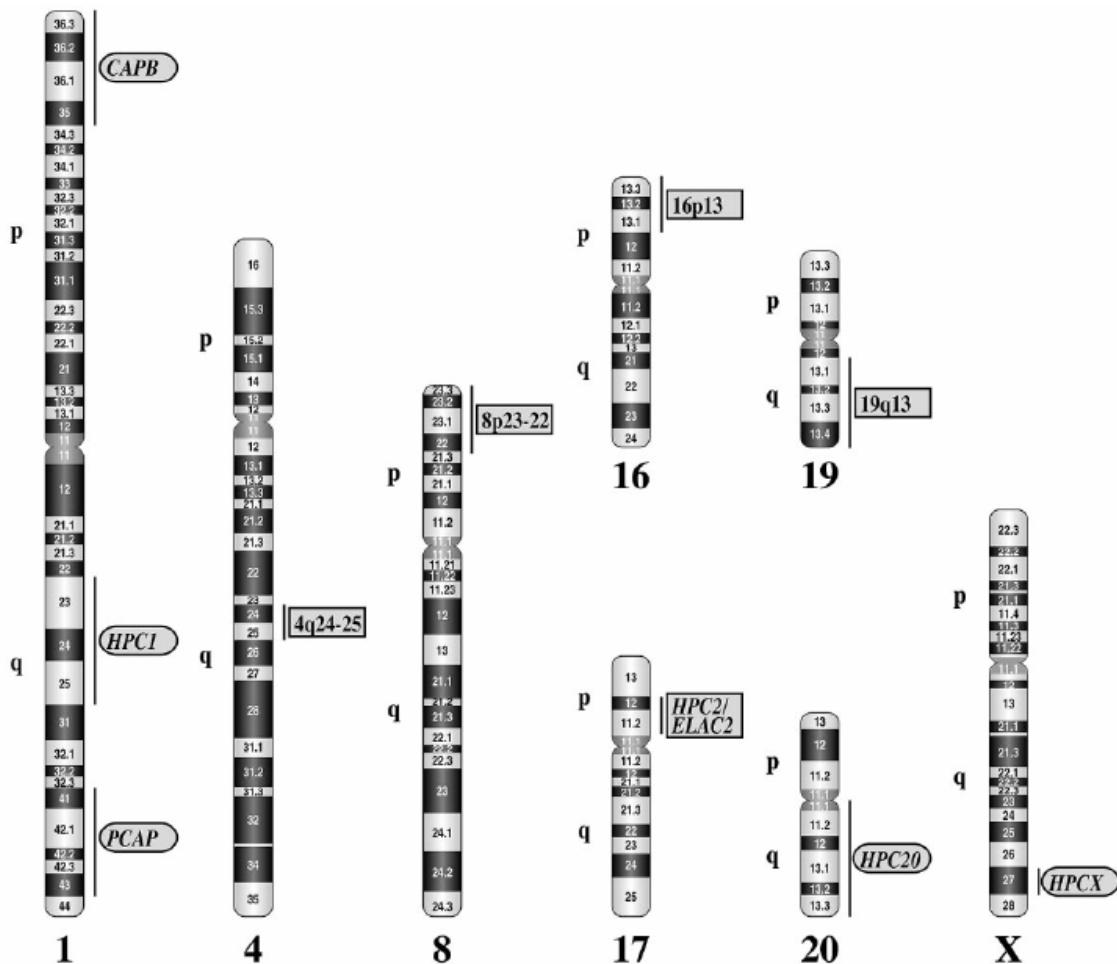

Figure 2.2:   Human Prostate Cancer Susceptibility Loci [9]

It has been more difficult than other common cancer, such as breast cancer, to associate candidate genes to this malignancy. Among these loci, HPC2 is the region of first discovered susceptibility gene, *ELAC2*. Other important loci are HPC1 harboring *RNASEL* gene, *MSR1* on chromosome 8p; which are among the best supported associations by linkage analysis. [9] Although, linkage analysis pointed important relations between certain genes and PC, results were mostly insignificant. This led to conclude influence of several but lower penetrance genetic variants. Apart from linkage analysis, other groups used epidemiologic and pathologic findings to identify susceptibility genes. Androgens have a fundamental role in prostate carcinoma, and to identify such genes, enzymes in androgen synthesis pathway would be compelling targets. Among these enzymes, 5-α-reductase, CYP11A1, LH-β, AR, CYP-17 and CYP-19 have attracted attention both as genetic markers and drug targets. [10]

2.1.6. Prostate Cancer Treatment

Several methods are employed for treating prostate cancer. Chemotherapy, radiation therapy, surgical procedures and hormonal therapy are among these options.

Chemotherapy and radiation therapy are severe measures for controlling malignant tumors. They involve use of cytotoxic drugs or ionizing radiation to kill cancer cells. Therefore, these methods may cause severe, even lethal, side effects.

There are two surgical procedures followed for PC treatment, prostatectomy and orchiectomy. Prostatectomy is the removal of all or a part of prostate gland, and used to remove malignant tissue. Testicles are the major producer of androgenic hormones and androgens act as growth factors for PC. Therefore, orchiectomy, removal of testicles, is another method of PC treatment. Both these surgical procedures face side effects varying from urinary tract problems to infertility.

Hormone therapy is another treatment method. Since androgens play a fundamental role in prostate cancer, replacement of these hormones with anti-androgens or antagonists of androgens are effective therapy methods. Controlling the synthesis of

these hormones over gonadorellin, which is a regulator of androgen synthesis, is yet another hormonal therapy method. [11]

Figure 2.3 describes the action logic of hormonal therapies. Anti-androgens are used to prevent actual androgens to interact with androgen receptors and prevent proliferation. Estrogens are used to prevent gonadorellin release; hence, positive regulation of gonadorellin on androgen synthesis is controlled. Gonadorellin analogs work on the same principle, and suggest a way of decreasing hormonal therapy side effects. However, both are peptides and can not be administered orally. More seriously, they can not inhibit androgen synthesis completely. Gonadorellin is effective on testicular androgen synthesis, which accounts for almost 80% to 90%. The rest is produced by adrenal glands and can not be affected by use of gonadorellin or analogs.
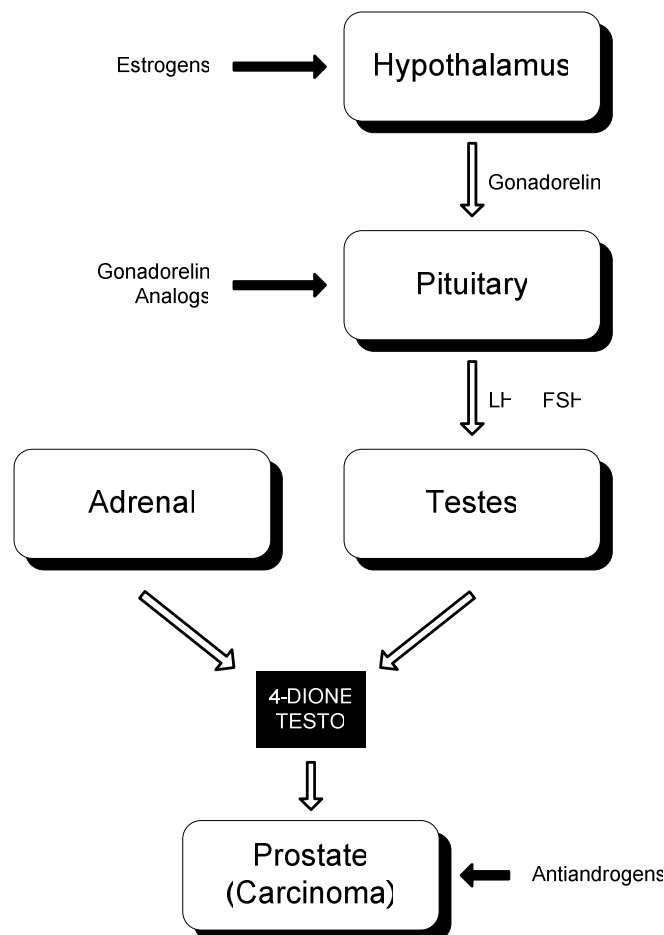

Figure 2.3:   Hormonal Treatment Strategies for Prostate Cancer (adopted from [11])

2.2.   Structure-Based Drug Design for Prostate Cancer

2.2.1. Drug Design

There are a variety of approaches in drug design ranging from purely experimental and integration of computational methods with experimental studies. High throughput screening, a purely experimental approach, test large number of molecules using combinatorial chemistry methods for protein binding and activity inhibition with robotics integrated systems. High experimental costs, small number of readily available chemical substances for testing, and possible interactions of molecule with other binding sites on target protein are among the main problems facing purely experimental methods. [12]

2.2.2. Structure-Based Drug Design

The use of X-Ray crystallography and NMR technology to determine protein structures created a new approach for drug design. Using protein structure data to test chemical structures stored in databases for binding to the active site of the protein is a powerful method. Structure-based or structure-guided drug design has been accepted as an established approach in pharmaceutical industry and academia. There are over 40 drugs generated using structure-based methods. The list of drugs developed with structure-based drug design approach includes drugs for hypertension, influenza, cancer and AIDS. [13]

Structure-based drug design is an iterative process that starts with target selection. [14] Once the target protein is found, structure should be determined by one of the three methods; X-ray crystallography, NMR, or computer-based modeling. Quality of the structure should be evaluated. One of the important requirements in the structure-based drug design is the structural analysis of the active site of the protein. Next step is the use of virtual screening to score chemical molecules from a database. As high score compounds are found, bioactivity of these molecules should be tested on target protein. Last two steps can be repeated many times to identify more candidate molecules or to

improve efficiency of previous candidates. [15] A slightly modified scheme of structure-based drug design is used in this thesis (Figure 2.4).
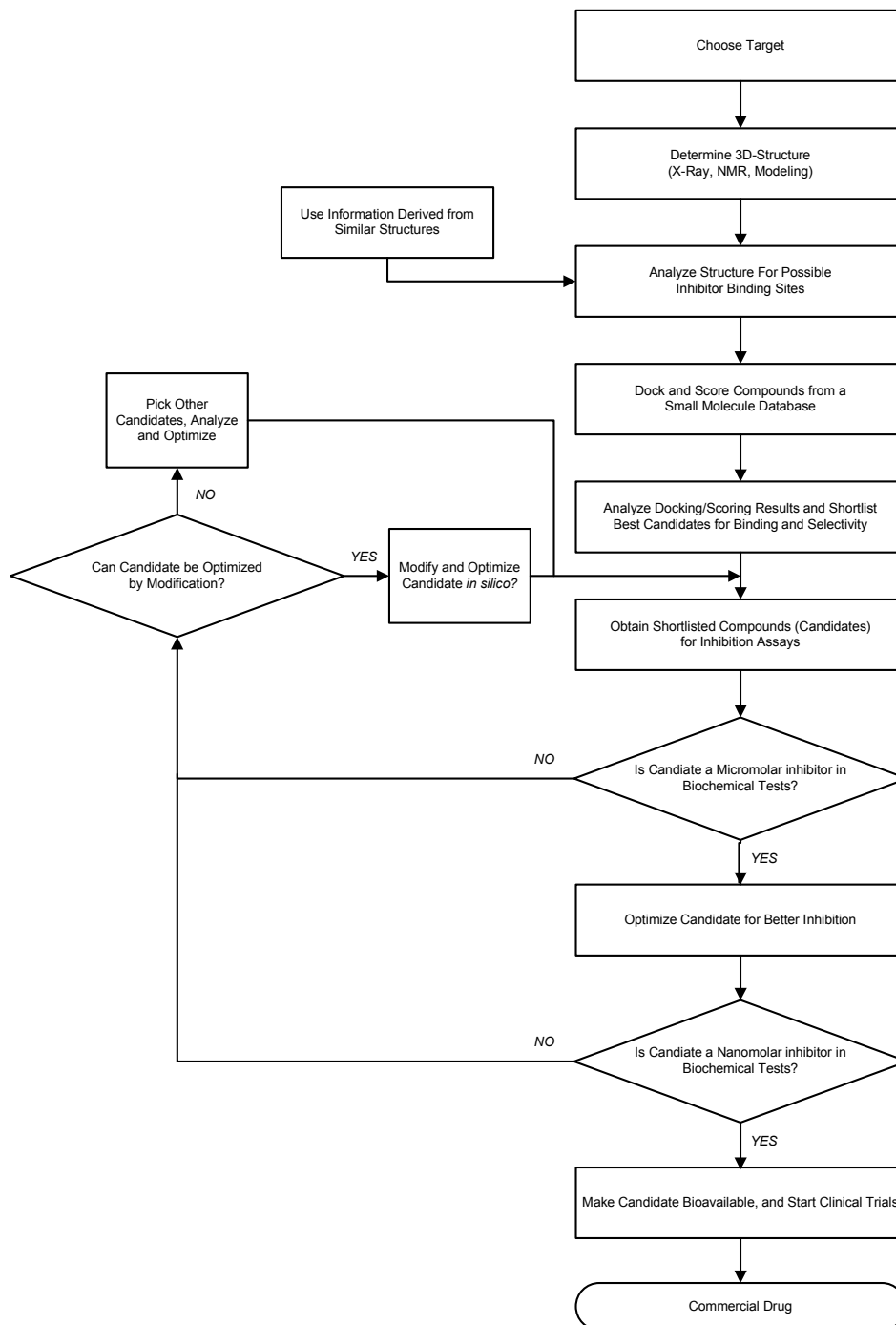

Figure 2.4:   Structure-Based Drug Design Scheme (adopted from [14])

2.2.3. Androgens and Target selection for Prostate Cancer

Androgens are steroid based hormones. Their main function is differentiation, maturation and regulation of sex specific organs. Androgens are growth factors of prostate cancer. Major androgens on this regard are testosterone (TESTO) and dihydrotestosterone (DHT). These molecules act on androgen receptor (AR) and AR regulates expression of several genes, leading to cell proliferation.

According the recent study 90% of the patients respond to androgen deprivation. [16] Therefore, suppressing androgen biosynthesis is an alternative PC treatment strategy. As explained in previous sections, using gonadorellin analogs or antagonists is not completely effective due to adrenal androgen release and tumor growth can not be fully blocked. If androgen synthesis can be inhibited, the side effect of chemotherapy or other methods can be avoided. [11] In order to find a suitable target protein, androgen synthesis pathway is analyzed. (Figure 2.5)

Testosterone (TESTO) and dihydrotestosterone (DHT) are the most important androgens regarding PC. [17] As shown in Figure 2.5, androgen biosynthesis occurs by participation of three tissues; adrenal glands, testis and prostate. Androgen synthesis pathway starts from cholesterol and then several enzymes are catalyzing reactions for formation of TESTO and DHT. CYP17 catalyzes conversion of progesterone to androstenedione (4-DIONE); 17-β HSD-3 which is intermediate key compound for TESTO; and, finally, 5-α reductase converts TESTO to more potent androgen DHT. The limiting step in the androgen biosynthesis is catalyzed by CYP17. Studies have indicated that blocking of CYP17 activity result in blocking in androgen biosynthesis. Therefore, development of new inhibitor with less side-effect against the CYP17 will prevent androgen biosynthesis and these inhibitors can be used as effective PC treatments. [11]
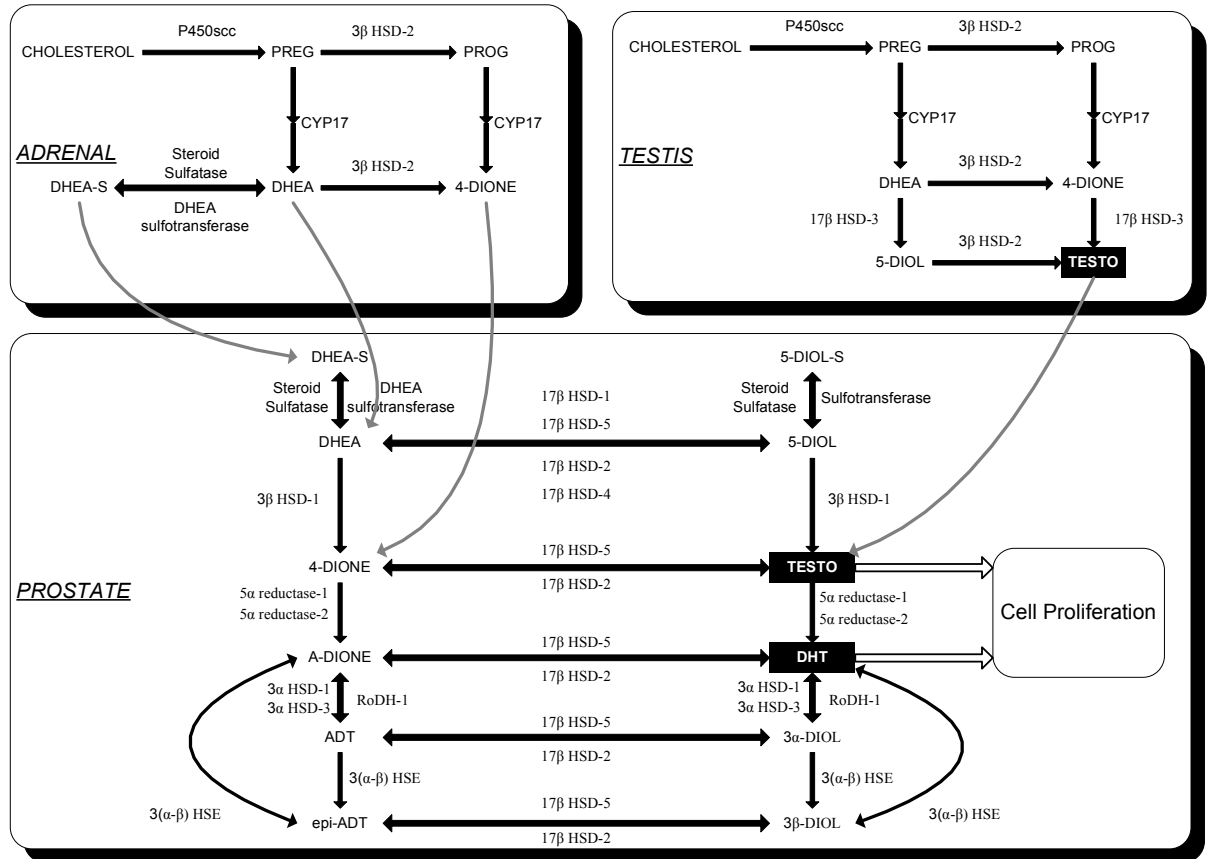
Figure 2.5:   Androgen Synthesis Pathway (adopted from [17])

2.2.4. CYP17: P450 monooxygenase 17-α Hydroxylase / 17-20 lyase

CYP17 is a microsomal enzyme and encoded by a single gene that is located at 10q24.3. [18] Although, this locus has not been previously reported in hereditary prostate cancer loci, two variants at the promoter of this gene have been studied for PC association. One of these alleles cause higher expression level; however, a significant association between PC and CYP17 promoter variants have not been reported.[10, 19]

CYP17 is found in the mammalian adrenal, the testes, and ovarian theca tissue. [20-25] In the testes and adrenal glands, the last step in the biosynthesis of testosterone involves two key reactions. These reactions act sequentially, and CYP17 is the enzyme responsible for catalyzing the two reactions. The first reaction is steroid 17α-

hydroxylation, in which an oxygen is inserted into C_{17}-H bond; and the second reaction is 17,20-lyase activity, in which 21-carbon 17 α-hydroxysteroids are cleaved to 19-carbon, 17-ketosteroids. (Figure 2.6) Pregnenolone (PREG) and progesterone (PROG) are good substrates for the 17α-hydroxylase reaction, but 17-OH-PREG is preferred over 17-OH-PROG for the 17,20-lyase reaction. [26]
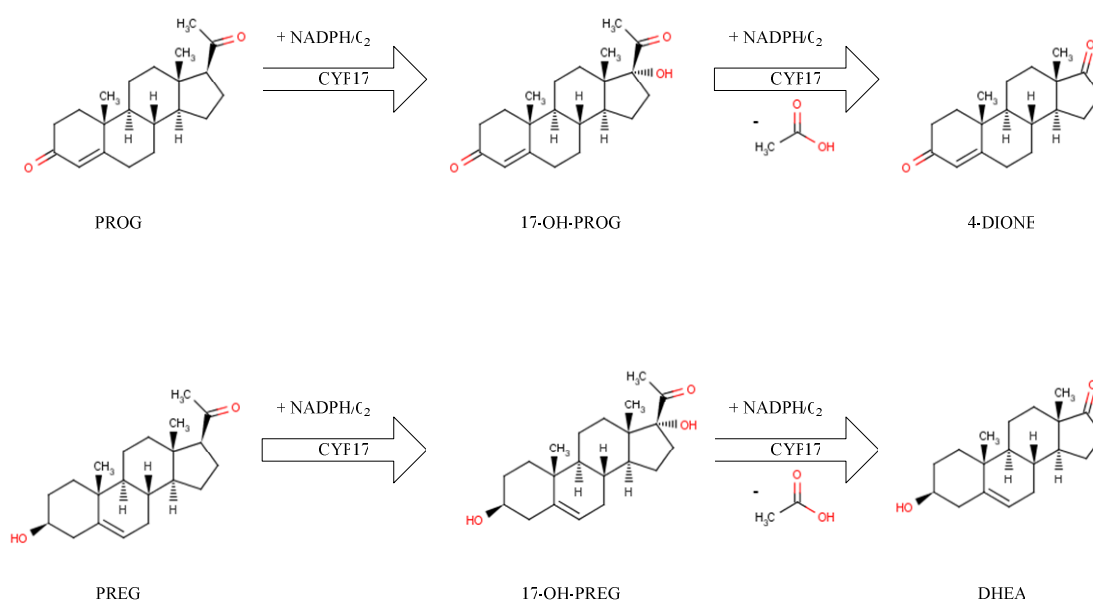
Figure 2.6:   17 α-Hydroxylation, 17-20 lyase activities of CYP17

2.2.5. CYP17: Structure

Enzyme structure is crucial for understanding the catalytic activities, substrate and reaction selectivity. Knowing the structure of CYP17 is mandatory for designing specific drugs to inhibit the catalytic activities of the enzyme. The crystal structure of CYP17 has not been solved yet. Since CYP17 is a microsomal P450, crystallization has its own difficulties. However, there are studies on computer-generated models of protein or active site. Laughton *et al*. [27] also build a model for CYP17, and Lin *et al*.

[28] modeled the active site of the protein. Both of these models were based on the crystal structure of P450cam, a class I P450. Lin *et al.* defines a bi-lobed substrate binding pocket. These studies were based on the crystal structures of class I P450s, those that use ferrodoxin intermediate as electron donor. However; class II (microsomal) P450s, like CYP17, modeled using class I P450s as templates have been suboptimal. The more recent model by Auchus *et al.* [29] (PDB ID: 2C17) is based on a class II P450 crystal structure, P450BMP. Therefore, the more recent model based on a class II P450 crystal structure, P450BMP, will be used in this study.

Modeled structure has a mono-lobed, hydrophobic binding pocket; meaning both reactions catalyzed by the enzyme occurs in one active site. The mono-lobed substrate binding pocket is defined with heme group, as the floor of the pocket; residues I288 to H321 which makes the I-helix and stands on one side of heme group; residues P368 to K374 and K91 to G95 forming the 4th and 5th strands of β-sheet 1, standing opposite the I-helix; I112; residues 365 to 367 that makes a loop after K-Helix; and finally V482 and V483 making up the top portion of the pocket. In this model, C_{17} of substrate steroid molecule would be above heme iron. Residues P368, V482, V483 and A302 would have hydrophobic interactions with C_{18} and C_{19}. Oxygen on C_3 of steroids would be making hydrogen bonds with residue G95. More specifically, Δ^4 (3 keto) steroids forms hydrogen bond with amide hydrogen of G95, and Δ^5 (3β-hydroxy) steroids form a hydrogen bond with the carbonyl oxygen, on MD simulations. All these residues form a smaller substrate binding pocket than the previous models, letting only the planar substrates like steroids to accommodate. Catalytic activity of the enzyme relies on a protein donor, residue T306. The hydroxyl group of this residue is placed across the heme ring from the D ring of steroid substrate. The place of T306 is homologous to other P450 crystal structures, and the hydroxyl proton at this location enables 17α-hydroxylase reaction by taking role in dioxygen protonation and O-O bond cleavage. [29]

2.2.6. CYP17: Inhibitor Studies

CYP17 is a well-recognized target for prostate cancer treatment, since selective inhibition of the enzyme exerts control over androgen synthesis. Therefore, many experimental and computational studies were reported on inhibitors of CYP17. First inhibitor designs used PREG and PROG as feeds since no drugs effective on CYP17 were reported. Due to use of substrates, first generation of designed compounds was steroid based molecules with various side chains attached to 17th carbon. Even though some of these compounds showed inhibitory effects, they were not used for the treatment of PC. The drawbacks of steroidal compounds were poor acid-stability, poor bioavailability, short half-life, first-pass effects and poor selectivity. [11] Ketoconazole is an antifungal agent known to reduce androgen levels in human, and has inhibitory effect on CYP17. However, it has been removed from use because it has toxic effects on liver and inhibition on other cytochrome enzymes. [30] Ideyemaya *et al.* worked on a non-steroidal inhibitor (YM116), which is able to inhibit both 17α- hydroxylase and 17-20 lyase reactions catalyzed by CYP17 with IC_{50} values lower than ketoconazole. This non-steroidal compound was 50 fold more effective in inhibiting 17-20 lyase reaction than 17α- hydroxylase reaction, which confirms selective inhibition potential of YM116. [31] Nnane *et al.* presented novel steroid-based inhibitors of CYP17. The IC_{50} values for five steroid-based compounds were determined for CYP17 and 5α- reductase. Molecules L-6 and L-26 showed more potent inhibition than ketoconazole. Despite their problems in bioavailability, these compounds were found to be promising as potential agents for reducing levels of testosterone and dehydro- testosterone in patients with androgen dependent diseases. [32] Effects of a number of thiazolidinediones were reported by Arlt *et al.* Thiazolidinedione and biguanide drugs, which are used to increase insulin sensitivity in type 2 diabetes, lower serum androgen concentrations in women with polycystic ovary syndrome. In order to determine if this is a secondary effect or direct control over androgen biosynthesis, these compounds were subjected to activity assays. Thiazolidinediones were found to exert inhibition on both reactions catalyzed by CYP17, but not biguanides. In another study, the effect of

cinnamic-acid based derivatives of thiazolidinediones on CYP17 was analyzed. Although these studies do not focus on the treatment of prostate cancer, some of the reported compounds have shown inhibition on both reactions catalyzed by CYP17. [33, 34] C-17-Heteroaryl steroidal compounds were rationally synthesized and tested for inhibitory and antitumor effects by Handratta *et al.* Some of these benzoazoles and pyrazines were found to be potent inhibitors of CYP17 as well as being antagonists of androgen receptors. [35] The work of Clement *et al.* uses steroid-based inhibitors and generates a pharmacophore model of human CYP17 inhibitors. [30] This model is used to retrieve hits from Maybridge, ACD and BioByteMasterFile chemical databases. Hartmann *et al.* also published a review on inhibition of CYP17 by steroidal and non-steroidal molecules as a method for androgen-dependent prostate cancer treatment. [11]

Chapter 3

Materials and Methods

3.1.   Computational Methods

3.1.1. Molecular Dynamics Simulation

Molecular dynamics (MD) is the computation of interaction of atoms in a molecular system by numerically integrating Newton's equations of motion. MD uses molecular mechanics methods to calculate time dependent motions of this system. Thereby structure and key properties, such as stability and binding between molecules, of the system can be studied. MD has become a tool for analyzing the dynamic characteristics of molecules.

The first simulations were done on systems containing ~500 atoms and the period was 10ps. With the advances in MD softwares and algorithms, the study of dynamics of large macromolecules, including biological systems such as proteins, nucleic acids (DNA, RNA), and membranes are feasible. Moreover, the simulation times are now 1000 times longer, where actual time required for these simulations are ~50 fold less. [36]

Molecular dynamics simulations serve biological sciences in several important means. One of these is the use of MD for revealing the effects of solvents and temperature on protein structure. Protein molecules interact with surrounding solvent molecules and these interactions play role in determining internal motions of proteins, as well as actual 3D structure. [37] Protein folding has long been an interest of researchers, combining experimental evidences and MD simulations led to a general understanding for this fundamental problem. [38] Since most functions in cells are done

in multi-subunit protein structures, protein-protein interactions have become another field for MD simulations.

There are a few MD software packages and force fields available to life sciences researchers. Among them, NAMD [39] is used with CHARMM [40] force field parameters. VMD [41] is a graphical user interface used to prepare and evaluate MD simulations.

NAMD is a molecular dynamics software that enables high-performance simulation of biological macromolecules in realistic environments. For this purpose, an efficient numerical integrator for Newtonian equations of motion is used. In order to control temperature and pressure statistical mechanics methods are employed. For high performance evaluation of electrostatistic forces, partial mesh Ewald is used.

As a function of time, the position of each atom is calculated by iteratively solving Newton's equation of motion,

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} = -\frac{\partial U}{\partial r_i}$$

(3.1)

If the potential energy is know, then with a given set of coordinates and initial velocities, force acting on each atom can be calculated. Hence, a new set will be obtained. Repeating these calculations will result in a trajectory for atoms through the changes in the system.

NAMD performs all-atom simulations, where model force field is used to estimate the interaction of each atom with the rest of the atoms in the system. A common energy function (3.2) is used in NAMD which has terms for bonded, angular, dihedral potentials, as well as terms for van der Waal's and Coulomb interactions:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{vdW} + U_{Coulomb}$$

(3.2)

The first three terms will define stretching   (3.3), bending   (3.4)   and   torsional (3.5) interactions:

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

(3.3)

$$U_{angle} = \sum_{angles=i} k_i^{angle} (\vartheta_i - \vartheta_{0i})^2$$

(3.4)

$$U_{dihedral} = \sum_{dihedral=i} \begin{cases} k_i^{dihedral}[1+\cos(n_i\Phi_i - \gamma_i)], n_i \neq 0 \\ k_i^{dihedral}(0_i - \gamma_i)^2, n=0 \end{cases}$$

(3.5)

Each covalent bond between a pair of atoms, each angle between adjacent bonds and each dihedral between pairs of atoms separated by exactly three covalent bonds, except improper dihedrals, will be used in these equations to calculate the potential created by the system over every single particle.

The later two terms are used for interactions between non-bonded atoms pairs:

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

(3.6)

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

(3.7)

(3.6) represents the van der Waal's forces approximated by Lennard-Jones 6-12 potential and (3.7) is for electrostatic interactions.

To overcome surface effects, boundaries are merged with a replica of the original system. A particle leaving from one boundary is replaced by the same particle entering from the opposite boundary; therefore, eliminating surface effects.

3.1.1.1.   MD Setup

Coordinate of the initial structure is obtained from Protein Data Bank (PDB code 2c17)[29]. Protonation states of histidine residues are done as they are reported in the

model structure. As the initial structure was a model, all hydrogen atoms were already present; therefore, there was no need to place hydrogen atoms by guessing. CYP17 is composed of 508 amino acids; however, model contains 454 amino acids. The reason for missing amino acids is the nature of protein and the template used to create the model. CYP17 is a microsomal enzyme, meaning that it has a membrane embedded domain. This membrane spanning domain is located on the N-terminus, beginning of the protein. Since it doesn't have an active role in protein's catalytic function, it is deleted from the model. Moreover, some residues in C-terminus, end of the protein, correspond to distorted residues of the template due to unclear diffraction pattern. To increase the overall quality, these residues are also deleted in the model built by Auchus *et. al.* [29]. VMD is used to prepare protein structure file (PSF) for NAMD simulation. Using *psfgen* package, atom and residue names are replaced with ones recognized by NAMD. Protein and heme moiety is initially treated as separate chains, and linked after *psfgen* made necessary replacements. Parameters for heme-cystine are set according to work of Collins *et. al.*[42]. VMD Solvate plug-in is used to solvated structure in a water-box with minimum 10 Å distance from any atom of protein to the boundary. Finally, system is neutralized with the addition of Na, or Cl ions with VMD Autoionize plug-in.

Equilibration of the system is done in multiple steps. First, 10000 steps minimization on non-backbone atoms and then 10000 steps for all atom are performed with no pressure control. Second, system is brought to physiological temperature (310 K) by 10 K increment steps where alpha-carbons and heme is restrained (restrain constraint is 1 kcal/(mol Å^2)). For each of these steps, 10 ps simulation is done with Langevin temperature set accordingly. Finally restrain on noted atoms are released by 0.25 kcal/(mol Å^2) restrain constraint at a time with 10 ps in between.

Simulations are done on Koç University High Performance Computing (HPC) Laboratory in parallel; using NAMD 2.6 and CHARMM27 force field parameters. Time-step for the simulation was 2 fs for calculation of bonded interactions, van der

Waals interactions (12 Å cut-off), and long-range electrostatic interactions with partical-mesh Ewald (PME). Constant pressure control is applied by Langevin dynamics coupled to all atoms with a 5 ps^{-1} damping coefficient. Constant pressure is held at 1 atm with a decay period of 100 fs and a damping time of 50 fs. Using this setup, a 10 ns simulation is obtained. Relevant configuration file can be found in supplementary section.

3.1.2. Molecular Docking

3.1.2.1. Protein-Small Molecule Docking

Molecular docking has become an indispensable part of structure-based drug design in the last decade. In the most general sense, docking is prediction of optimal energy and spatial configuration between two molecules. Therefore, by docking, the optimum configuration that maximizes the interaction between protein and drug is searched.

In almost all docking approaches, protein molecule is set as a rigid body, until recently Autodock4 is released as flexible protein-flexible ligand docking algorithm. Excluding Autodock4, protein-ligand docking can be classified into two classes, where the ligand molecule is rigid or flexible. Rigid-ligand docking is faster, which is advantageous for database screens. Since the molecule is kept rigid, molecule's all conformational space can not be sampled with docking. Main principle behind ranking of this approach is geometric, meaning volume and surface complementarities. Therefore, algorithms are built on shape matching between surface or clefts of protein and surface of ligand. Examples of rigid docking software are GOLD [43], Dock [44], and FRED [45]. On the other hand, flexible ligand docking was less suitable for 3D-database screens due to longer processor times. Unlike rigid-ligand methods multiple-conformations of ligand molecule is used during docking or generated as the docking progresses. Selection of hit molecules is based on either geometrical constraints or interaction energies between protein and ligand. Docking poses and ligand conformations can be generated by several methods like simulated annealing, fragment

building or genetic algorithms (GA). GOLD, FlexX [46], MOE [47], Dock and Autodock are examples of flexible-ligand approaches.

Main problems of docking are identification of interaction site for a ligand, large search space of ligand conformations and binding site, and ranking of docked molecules. The problem can be overcome with prior knowledge of protein. This information can be used to focus on active site of protein, if the aim is to design an inhibitor; target interaction sites, if protein-protein interactions are to be disrupted; or other effector-sites to change regulation of a proteins' function. Conformational space of a ligand molecule can be very small or extremely large depending on the number of torsional freedom. For instance, a steroid based molecule without long side chains has restricted flexibility, hence small conformational space; where a similar molecular weight hydrocarbon molecule can assume various conformations. Since it is not possible to evaluate all samples in this space, GA is employed for faster and accurate results. Finally, rankings are done based on geometrical and energetic constraints. Ligands or conformations with better geometrical fit and lower estimated binding energy are ranked higher.

3.1.2.2. AUTODOCK

AUTODOCK is a freely distributed software developed for predicting macromolecule-ligand complexes, where macromolecule is rigid and ligand is flexible. AUTODOCK's working process can be basically described as docking simulation and scoring.

Docking simulation is done as a search for the best ways to put protein and ligand together. AUTODOCK offers several alternatives for stochastic search methods; such as Monte Carlo Simulated Annealing, Genetic Algorithm and Lamarckian-Genetic Algorithm (LGA). Since LGA offers the best for high degrees of freedom and is suggested by the producers, this search method will be used in the study.

LGA is a hybrid of static (local) and stochastic (global) methods. Evolutionary GA pursuit the understanding developed by natural genetics and biological evolution. Crossovers between parents and mutations are keys to produce favorable traits in children. In this regard, a ligand's state corresponds to genotype and atomic coordinates correspond to phenotype. Each variable required to define a state, such as torsional angle, or position with respect to protein, is translated as a gene. Finally, fitness is of a child is decided by its interaction energy. GA process starts by generating a random population of individuals (50-250). Pairs of individuals are selected for crossover and generate children, during which mutations can also occur. Then, fitness of each child is calculated and genotypes giving rise to favorable phenotypes are kept for next generation. In LGA, phenotypic advantages acquired during lifetime are returned to genotype, like Jean Batiste de Lamarck's discredited idea. These phenotypic advantages are discovered by local search algorithm. This hybrid method is found to be the most efficient one handling more degrees of freedom [48].

Calculation of interaction energies for ligand conformation and every atom position demands high amount of computational resources. In order to overcome this problem, AUTODOCK uses AUTOGRID component. This tool generates three dimensional lattice of points. One such grid map is generated for each atom type in ligand, named as *atomic affinity grid map*. Each point is assigned with the relevant non-bonded interaction potential with protein. This "look-up table" eliminates the need for most computation. Also, one map is generated for electrostatic potentials. Macromolecule is represented by these maps for AUTODOCK.

3.1.2.3.   Docking Setup

Running an AUTODOCK process requires 4 input files; PDBQS file for macromolecule, PDBQ file for ligand molecule, GPF file for grid parameters and DPF file for docking parameters. AUTODOCK Tools (ADT) is a GUI capable of producing these files.

Macromolecule file is prepared by first adding polar hydrogen atoms. However, CYP17 molecule after MD simulation has all hydrogen atoms. Therefore, non-polar hydrogen atoms are merged in this particular case. AUTODOCK permits use of two special atom types for macromolecules. These are required when there is a metal atom in the structure. In the case of proteins heme groups, two special atoms are required. Four nitrogen atoms in heme will be treated as special atoms, as well as iron atom. Pyrrole nitrogen atoms in heme rings are not capable of forming hydrogen bonding. Therefore, they have to be treated separately. Finally, ADT is used to add Kollman charges to structure, and these are recorded in PSBQS file. PDBQS files are prepared for 10 protein structure files extracted from the last ns of MD simulation, each with 01.ns in between.

Ligands are prepared similar to macromolecule file. First non-polar hydrogen atoms are merged. If the ligand is a molecule composed of amino acids, Kollman charges are added; otherwise, Gasteiger charges are added. In ligand, carbons are treated as aromatic, planer, or non-planer. ADT, also, defines carbons as planer or non-planer by the angle between respective carbon atoms. Finally, bonds with torsional freedom are marked and recorded to PDBQ file. Ligand files are obtained from a commercial compound database, AMBINTER. These files are filtered according to their molecular weights: molecules heavier than 300 g/mol are used in virtual screening.

Grid files are based on GPF files. On this file, grid center, grid size as points, spacing between grid points, and the number of grid files based on atom types are defined as parameters. In this thesis, two sets of grid files are used. The first set is a designed for a relatively coarse docking, and second is designed for fine docking. For virtual screening, low resolution grid maps are created to ease docking simulation. For this purpose, map size is defined as 52x52x52, spacing is kept as default 0.375 Å. This setting covers entire active site of the protein. Since, virtual docking is performed over several thousands of compounds, grid maps are generated for atom types present in all drug-like molecules. Fine docking setup has an increased resolution to allow study of

ligand-protein interactions in detail. Map parameters are defined as 126x126x126 grid points, and 0.150 Å spacing. As in previous setting, maps are generated for all required atom types. For both setups, Heme iron occupies the center of grid points. Samples of these GPF files can be found in Supplementary.

DPF contains parameters for docking algorithm and necessary variables for this method. Since LGA is used, population size, number of generations, number of runs, crossover rate, mutation rate and number of evaluations should be defined here. As in the case of grids, a coarse docking setup will be preferred for virtual screening. Population size is 50, number of generations is 27000, crossover rate is 0.8, mutation rate is 0.02, number of runs was 10, and number of evaluations was 1000000. For fine docking, search parameters were improved. Population size is 250, number of generations is 50000, crossover rate is 0.8, mutation rate is 0.02, number of runs was 100-256, and number of evaluations was 15000000. Samples of these DPF files can be found in Supplementary.

3.2.   Experimental Methods

All protocols and solutions used in this section are covered in detail in Supplementary section.

3.2.1. Construction of human CYP17 cDNA into Mammalian Expression Plasmids

CYP17 gene was kindly provided by Dr. Richard J. Auchus in a pcDNA3.1 plasmid. CYP17 gene obtained from this plasmid is introduced pcDNA4 plasmids. CYP17 gene is amplified from pcDNA3.1 plasmid by PCR using GTATTAATACTCGAGCGGCCGCCACTGGGTGCTACCCTCAGCCTGGGC and GCCCAGGCTGAGGGTAGCACCCAGTGGCGGCCGCTCGAGTATTAATAC primers which contains unique restriction sites for EcoRI and Not I is performed. Both pcDNA4 plasmid and CYP17 PCR product are subjected to EcoRI and NotI digestion. Digested DNA strands are gel purified before ligation to get rid of small oligo-

nucleotides. Then ligation and transformation are performed with conditions described in supplementary. Chemically prepared *E. coli* DH5α competent cells are used for cloning to obtain large quantities of the plasmid. Transformed bacteria are plated on ampicillin selective media. Random colonies are selected for plasmid DNA isolation by alkali lysis method. Isolated DNA is visualized with agarose gel electrophoresis after relevant restriction digestions to confirm correct constructs. For excessive plasmid DNA isolations, maxi-prep is performed for each construct.

3.2.1.1.   Mammalian Cell Protein Expression

Vectors expressing human CYP17 were transfected into human embryonic kidney cell line (HEK-293T) by the calcium-phosphate precipitation method (see supplementary info). Cells were initially plated at $5x10^4$ per well in 6 well Falcon flask and grown in Dulbecco's modified Egale2s medium supplemented with %10 (v/v) fatal calf serum at 37 ^{o}C under %95 air, 5% CO_2 humidified atmosphere. Brifly transfection is carried out as follow: $CaCl_2$ and pcDNA4-cyp14 plasmid DNA is mixed. 2XHEBS solution is added drop-wise to the mixture and air is bubbled into the mixture. This mixture is let stand for 8 minutes at room temperature. Plate with HEK-293T cells are covered with this solution and incubated for 16 hours. Next day medium was replaced with fresh DMEM containg %10 fatal serum. Celss are further incubated for two days to maximize the expression of human CYP17 protein. Then, cells are scraped-off and collected with centrifugation at 200xg for 10 minutes at 4°C for expression analysis. Integrity and expression level of CYP protein are analyzed by both 12% SDS-PAGE and western-blot analysis using anti-His antibody.

3.2.2. Calculation of IC50 Values of Drugs

Activity of CYP17 was performed using transfected HEK-293T by Acetic acid releasing assay (AARA). After the transfection of pcDNA4-cyp14 into HEK293T in 6 well plate by calcium chloride precipitation method, various amount of candidate drugs were added directly on the cells in the presences of the radioactively labeled H^3 labeled

substrate 17-OH Pregnenolone. Cells were further incubated at 37°C for 5 hours. To remove excess substrate and concentrate acetic acid in medium chloroform extraction is carried out. Aqueous phase is collected and mixed with charcoal suspension. After centrifugation at 1500xg, for 15 minutes, at 4°C; aqueous phase containing radio-labeled acetic acid is collected and radioactivity is measured with Liquid Scintillation Analyzer.

3.2.3. Toxicity Assays of Drugs on Mammalian Cells

HeLa cells are used for toxicity assays without transfection. Cells are plated in 96-well plates, each receiving ~5000 cells, and let stand in incubator for 1 hour. After cell attachment, either test compounds, solvents, or DMSO is added at appropriate amounts to wells. Controls received none of these compound, negative controls did not receive HEK-293T cell. Plates are incubated 24,hrs with test compounds and CellTiter-Glo® Luminescent Cell Viability Assay is performed as described on product manual. Plates are removed from incubation and equilibrated to room temperature for 30 minutes. Kit reagent is added and mixed for 2 minutes. Plates are incubated at room temperature for 10 minutes and luminescence is recorded.

Chapter 4

Results and Discussion

4.1.   Analysis of Molecular Dynamics Simulation

MD simulation analysis is done using VMD and its plug-ins. As described in the methods chapter, equilibration system is sampled every 2 ps and frames are used for both root mean square deviation (RMSD) and deformation analysis.
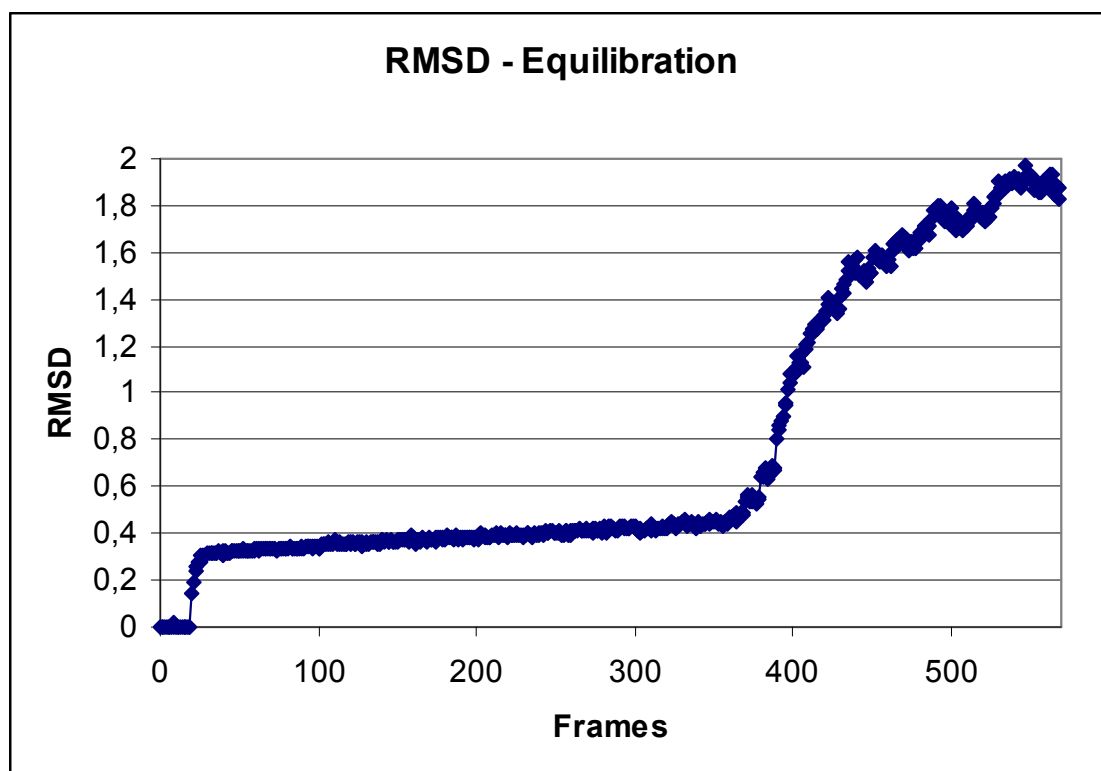

Figure 4.1:   RMSD – Equilibration

On Figure 4.1, RMSD for equilibration stage of the protein is plotted. Original modal structure is used as a reference. On this graph, first flat area corresponds to minimization of the protein where only hydrogen atoms are let to position themselves to best orientation. On the second flat region, system is gradually heated to physiological temperature where there is restraint on alpha carbons. Finally on the third part, where the restraint on protein is lifted slowly, RMSD increases at most to 1.9. CYP17 shows an expected pattern of solvation and relaxation during equilibration stage and no deformations are observed.
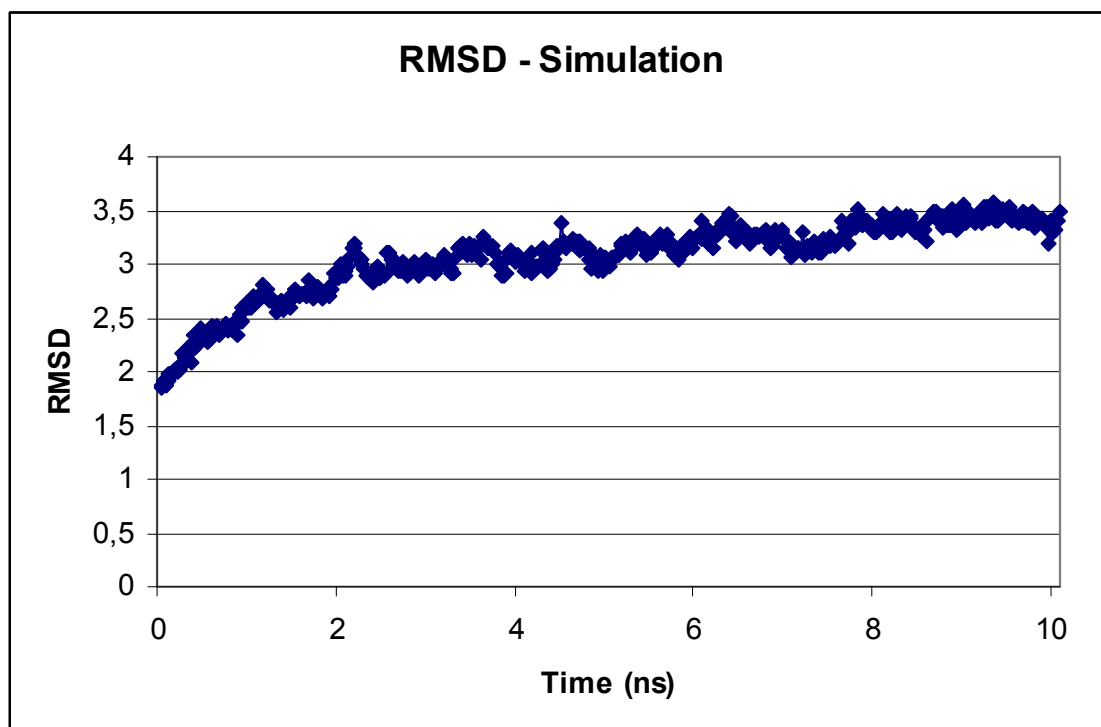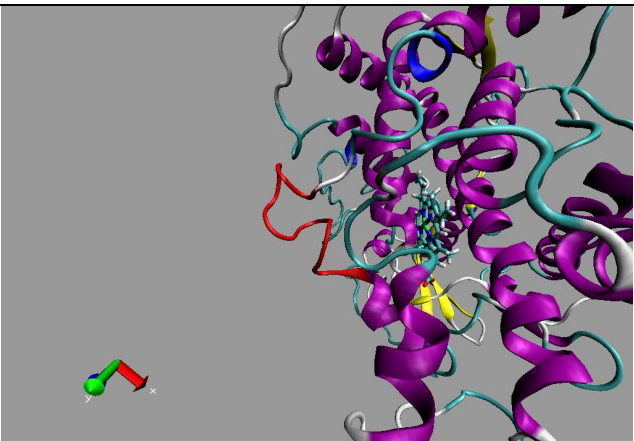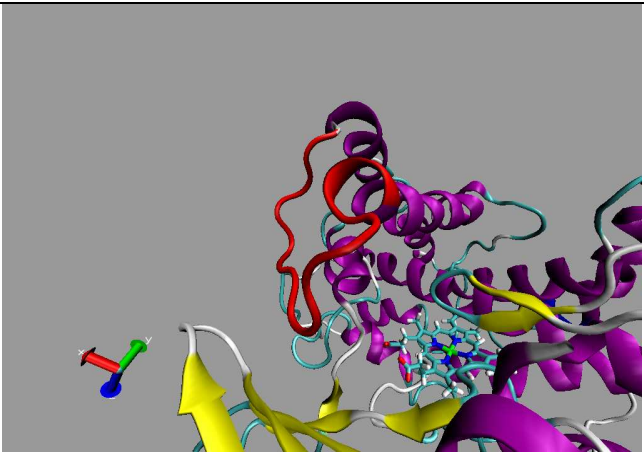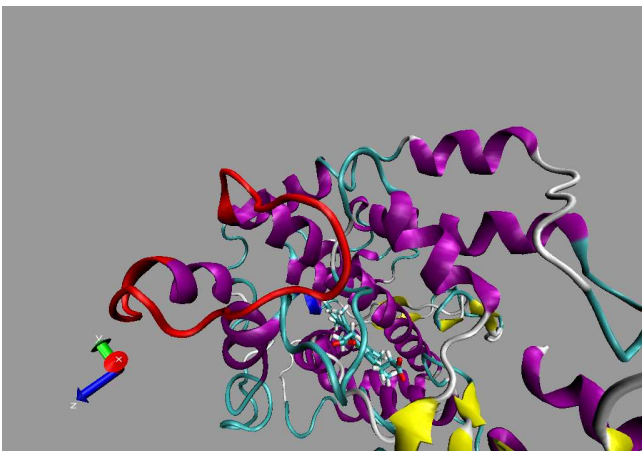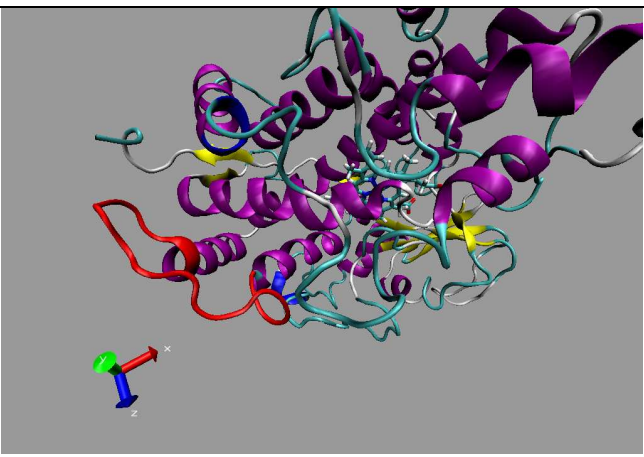

Figure 4.2:   RMSD – Simulation

On Figure 4.2, RMSD is plotted through simulation. As in RMSD for equilibration graph, initial modal structure is used as reference to compare generated structures. After all the restraints are released, RMSD continues to increase in the first 2 ns of the

simulation. For the rest of the 10 ns simulation, graph shows smaller deviations. This behavior of RMSD is an indication of stabilized protein in simulation. When RMSD is analyzed on residue level, it can be observed that there are regions of high and low deviations. High RMSD spots are concentrated in regions where there is no strict structure; in other words, where loops are located. Those regions are shown with their RMSD values and respective figures on Table 4.1. As shown on figures, all of these loops are on the surface of the protein. Since loops do not have regular hydrogen bonding pattern like alpha-helices or beta-sheets, they are not as stable. On the other hand, none of the alpha-helices or beta-sheets coincides with high RMSD spots (Table 4.2). Most importantly, the I-helix passing through the core of the protein and involved in active site is in a low RMSD region. Since this helix bears the catalytically active amino acid, it is crucial to have it in a stable form. MD simulation performed on CYP17 has proved the stability of active site.

Table 4.1:   High RMSD Regions

Residue / RMSD	Figure
83 2,838193 84 2,478919 85 3,304434 86 3,913952 87 5,029184 88 4,144932 89 4,803977 90 4,356662 91 3,92907 92 4,54816 93 4,746042 94 3,638751 95 4,156436	

164	3,446893	
165	3,516924	
166	5,562152	
167	7,028357	
168	5,218726	
169	6,144452	
170	4,568606	
171	4,446659	
172	2,99358	
173	2,557942	
174	2,110557	
175	2,597281	
176	2,438339	
177	3,552754	
178	2,833614	
224	3,661642	
225	4,082884	
226	4,444715	
227	5,989609	
228	6,446607	
229	6,281299	
230	6,195647	
231	6,742884	
232	7,002714	
233	7,471719	
234	7,040656	
235	6,728924	
236	4,078407	
237	4,603264	
238	3,655244	
239	5,318888	
240	3,128047	
241	2,582797	
289	2,096364	
290	3,532664	
291	7,820587	
292	3,44036	
293	4,206389	
294	4,754384	
295	4,947677	
296	5,181668	
297	5,28208	
298	4,82981	
299	3,480607	
300	3,585586	
301	4,002796	
302	3,747739	
303	2,259358	

328	3,810604	
329	4,813509	
330	3,276499	
331	3,630785	
332	3,689837	
333	3,102527	
334	4,258818	
335	4,197437	
336	3,781757	
337	4,502891	
338	2,985559	
339	2,570734	
340	2,83175	
341	4,210004	
374	4,792959	
375	6,062421	
376	5,301563	
377	5,847399	
378	3,86126	
379	4,111373	
380	3,099845	
381	2,941656	
382	4,101361	
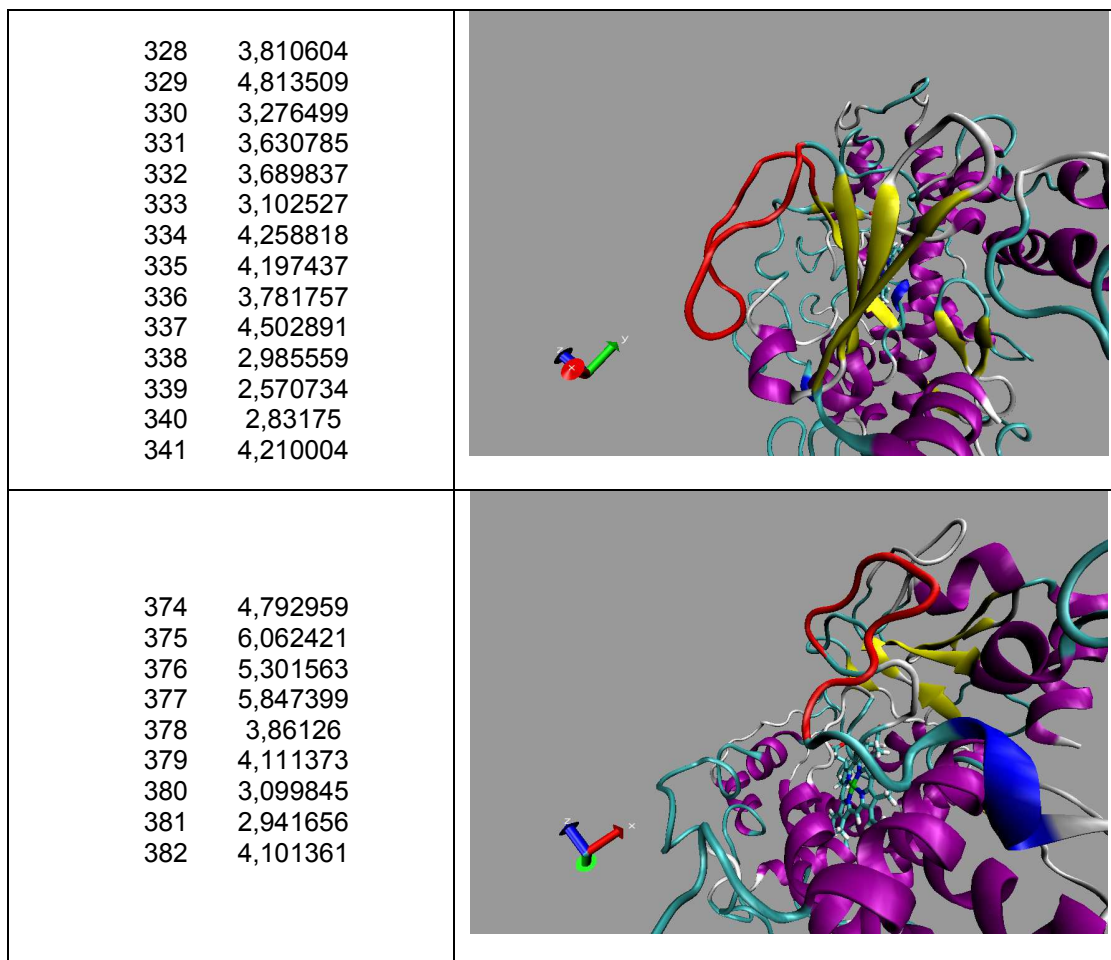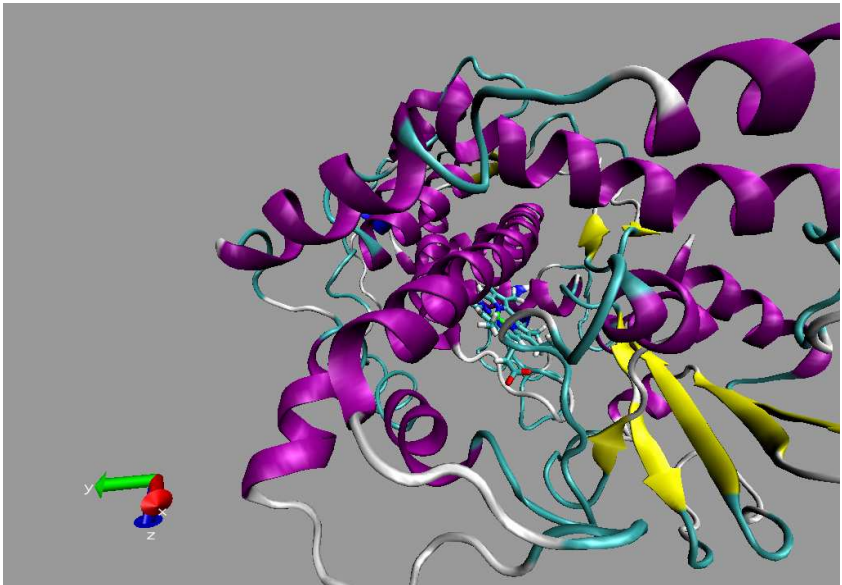
Table 4.2: RMSD per Residue - I-Helix

Residue / RMSD	Figure
242 2,383754	
243 4,079184	
244 2,0106	
245 2,04475	
246 3,429266	
247 1,872817	
248 1,035363	
249 2,858492	
250 1,586588	
251 2,764447	
252 2,178634	
253 2,409712	
254 0,794263	
255 1,34328	
256 1,404639	
257 2,259303	
258 2,201544	
259 1,666695	
260 1,739438	
261 2,029294	
262 1,939421	
263 1,456148	
264 1,917616	
265 2,411731	
266 1,616586	
267 1,167242	
268 1,978954	
269 1,63372	
270 1,899782	
271 1,563125	
272 2,151814	
273 2,113166	

4.2.   Analysis of Substrates

In order to validate the model, substrates of CYP17 should be docked in conformations suitable for reactions to take place. Detailed docking settings are used for evaluating the model with substrates. Docking conformations are shown on Table 4.3:

Table 4.3:   Docking for Substrates of CYP17
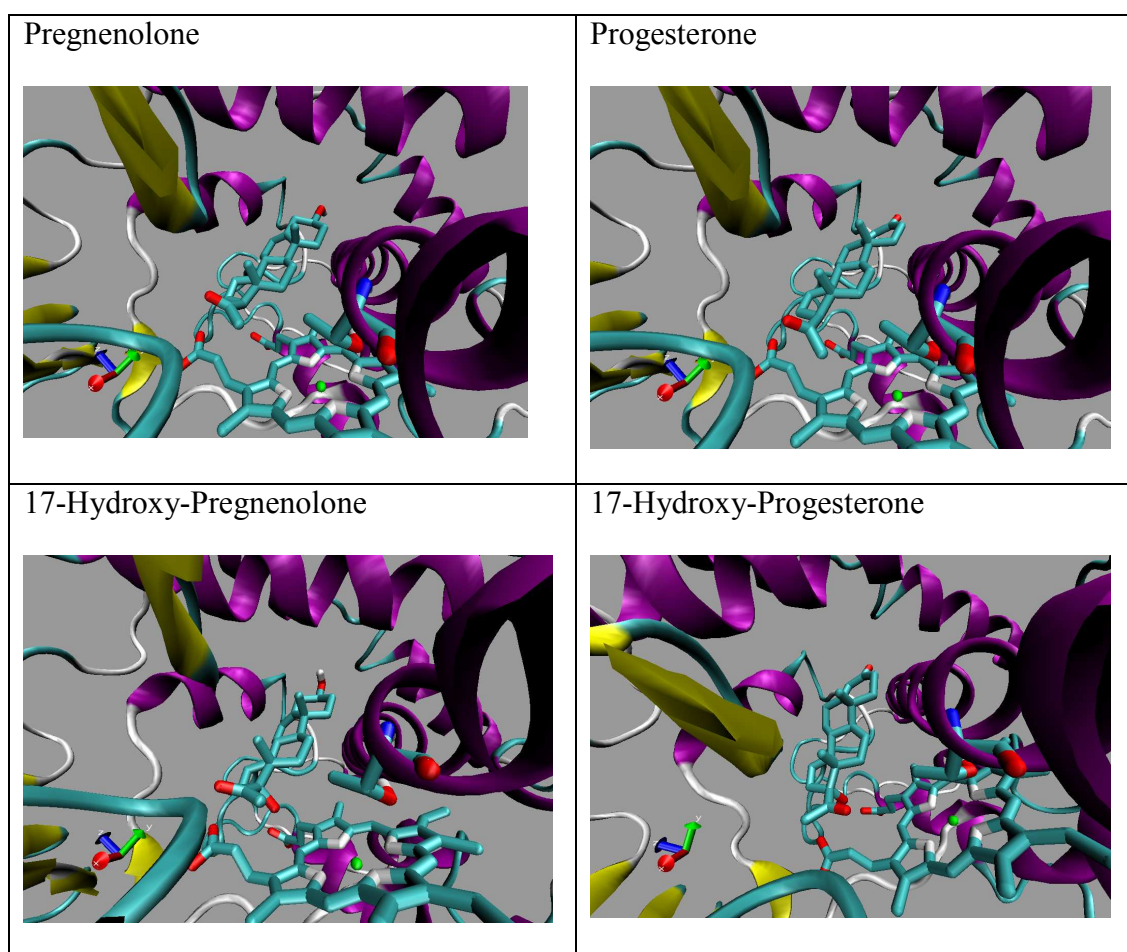
Docking positions are very similar to each other, since pregnenolone and progesterone have a common backbone. For pregnenolone and progesterone, the most important condition is accessibility of 17th carbon for hydroxylation. On the docking poses, both molecules have 17th carbon oriented towards heme and catalytically active threonine

residue. On the other hand, for 17-hydroxylated pregnenolone and progesterone, added oxygen atom should be well oriented for lyase reaction. On docking poses, both 17-hydroxy-pregnenolone and 17-hydroxy-progesterone have added hydroxyl residues accessible to heme and threonine residue. Docking energies for substrate molecules are given on Table 4.4.

Table 4.4:   Docking Energies for CYP17 Substrates

Substrate	Docking Energy	Substrate	Docking Energy
Pregnenolone	-10,12 kcal/mol	Progesterone	-10,39 kcal/mol
17-Hydroxy-Pregnenolone	-9,88 kcal/mol	17-Hydroxy-Progesterone	-10,42 kcal/mol

Docking energies for all substrate molecules are in the range of -10 kcal/mol. These results indicate a strong affinity of substrates towards CYP17. Docking energies obtained for substrates will also serve as a cut-off value for discovering potential competitive inhibitors in virtual screening.

4.3. Virtual Screening

In virtual screening phase more than 600,000 compounds in virtual library is considered. However, among these molecules, there was a large group having very small chance of inhibiting CYP17 due to size constraints. Since CYP17 has a rather small active site, it is not accessible for large molecules or molecules with many side chains. It is also found that most effective molecules had molecular weights around 300 g/mol [11]. Therefore, we applied a molecular weight, conserving compounds with molecular weights lower than 350 g/mol. When this constraint is used to filter the library, computational demand of virtual screening was eased greatly, reducing the number of compounds to be scanned by almost ten fold. A list for top 20 compounds with best docking and binding energies are on Table 4.5.

Table 4.5:   Energies for Top Scoring Molecules in Virtual Screening

Molecules	Binding (kcal/mol)	Docking (kcal/mol)	Molecules	Binding (kcal/mol)	Docking (kcal/mol)
5155	-11,67	-11,95	20543	-8,65	-15,08
3158	-11,48	-12,42	4216	-8,4	-14,06
24530	-11,41	-12,49	7435	-7,85	-13,94
16990	-11,4	-11,82	26310	-8,12	-13,89
6319	-11,37	-11,35	21614	-7,79	-13,7
10366	-11,34	-12,34	4603	-7,53	-13,7
15304	-11,25	-12,21	11461	-8,05	-13,63
14381	-11,25	-10,55	13767	-8,53	-13,58
20106	-11,24	-11,24	21660	-6,86	-13,47
21703	-11,21	-12,05	26579	-7,31	-13,39
17686	-11,16	-11,91	11311	-7,95	-13,33
14999	-11,14	-12,16	354	-7,52	-13,27
19096	-11,13	-12,05	14676	-8,15	-13,23
741	-11,13	-11,79	25356	-8,08	-13,18
23121	-11,12	-11,55	23167	-6,91	-13,14
19433	-11,09	-12,09	24707	-8,46	-13,13
28194	-11,02	-11,83	27983	-10,07	-12,99
157	-11,02	-11,4	13119	-10,68	-12,97
2350	-11,01	-11,82	2158	-9,49	-12,95
18295	-11,01	-11,16	20845	-8,2	-12,94

Compounds with binding and docking energies shown on Table 4.5 are strong candidates for inhibition of CYP17. However, among these compounds, several should be eliminated due to the following reasons:

- Docking far form active site

- Having long hydrocarbon chain in structure, high torsional degree of freedom

- Lack of a reactive species, such as nitrogen, or oxygen atoms, close to heme iron

Autodock gives the best energy conformation of ligand molecule in a given grid. However, this conformation is not necessarily on active site. Some molecules do not have suitable characteristics to interact with active site fruitfully. Those compounds may have good docking results, but their effect of inhibition can not be predicted based on docking results.
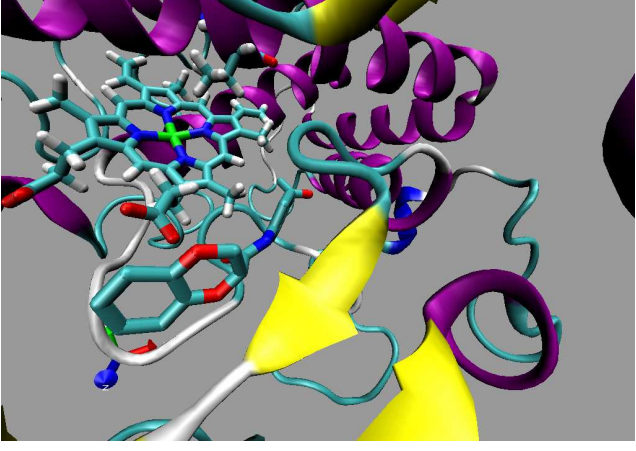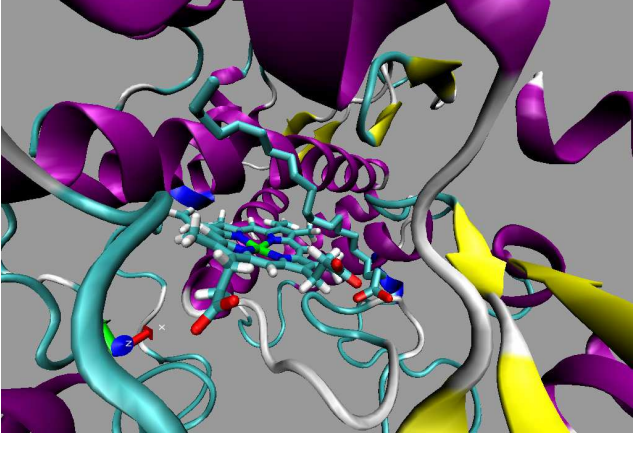
Hydrocarbons are very flexible and highly hydrophobic molecules. For a compound of molecular weight 300, torsional degrees of freedom can be as high as 20. Due to this flexibility, these molecules can assume almost any shape and, therefore, adapt to any active site. However, flexibility comes with a cost of conformational instability and torsional stress for compact conformations. These factors reduce docking scores of hydrocarbons and they should be low ranked. Unfortunately, high hydrophobicity of CYP17 active site gives these molecules better interaction energies, which increases their ranks, although they have no potential to be specific inhibitors.

Reactive species such as oxygen or nitrogen atoms give advantage to molecules to compounds in library. It is suggested to have such species in candidates to exert covalent modification on protein. [11] Heme and catalytic threonine reacts with such atoms in close proximity leading to better, even irreversible, inhibition.

Results are evaluated on the basis of above conditions. Several compounds are eliminated from the top list, some of which are sampled on Table 4.6. On the first sample, molecule has scored relatively low docking and binding energies. Compound is small, which makes it suitable for delivery to active site. However docking position of this molecule does not occupy active site. This molecule is not close to neither heme iron atom, nor the reactive residue. Therefore, it is not selected for further analysis. On the second example, subject molecule scored the best docking energy in virtual screening set. Molecule, also, passes over the entire active site. Since active site is occupied, molecule has a good chance of inhibiting CYP17 activity competitively. However, this is hydrocarbon derivative, a fatty acid, which has very high torsional degrees of freedom. Compound is very flexible, which enables it to assume many

conformations. One of these confirmations was fitted in the active site and resulted in a good docking energy. However, due to its flexibility, compound is not valuable as a specific inhibitor.

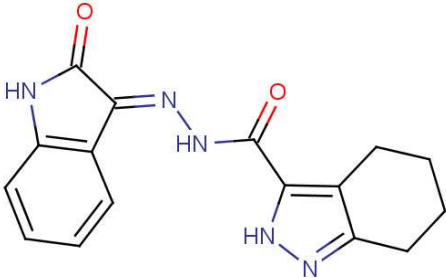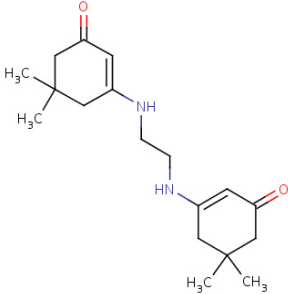Table 4.6:   Samples of Rejected Compounds

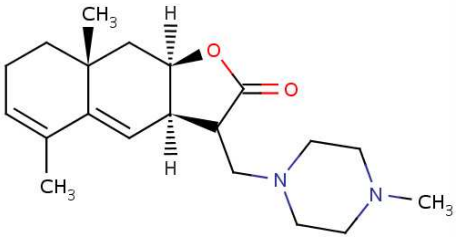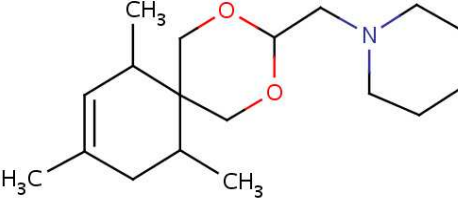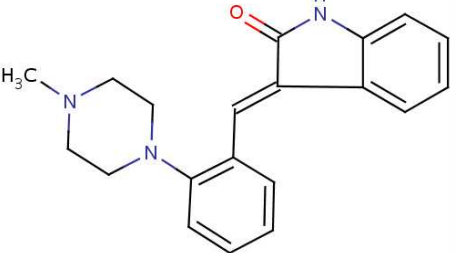| Docking Energy: -11,59 kcal/mol Binding Energy: -10,4 kcal/mol Reason for Rejection: Docking Position far from catalytically active residues	
Docking Energy: -15,08 kcal/mol Binding Energy: -8,64 kcal/mol Reason for Rejection: High torsional freedom	

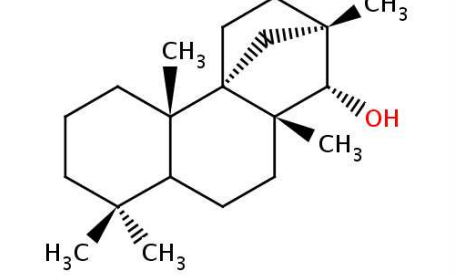Molecules with long hydrocarbon chains score very low docking energies as shown with the sample on Table 4.6. If selection of molecules for detailed analysis was only based on docking energy, many of those compounds had to be evaluated. However, when binding energies were involved in decision, there is a clear distinction. None of the compounds with long carbon chains has scored binding energies as good as
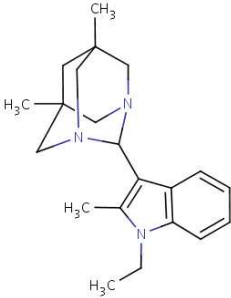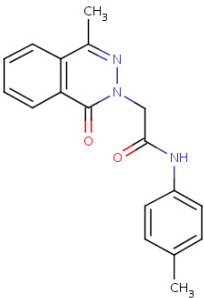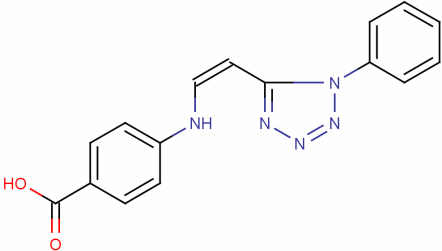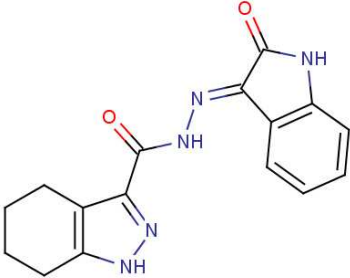
-10 kcal/mol. Therefore, a cut-off value of -10 kcal/mol is used as a filter for virtual screening results. This filter resulted in reduction of results to 400.
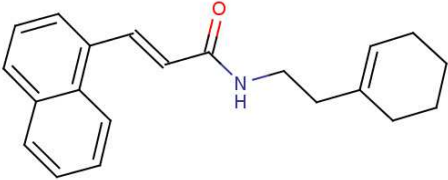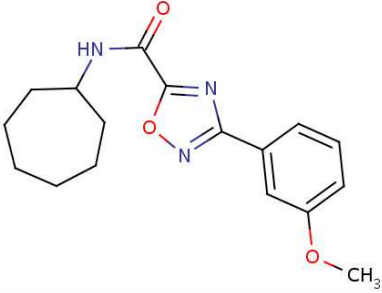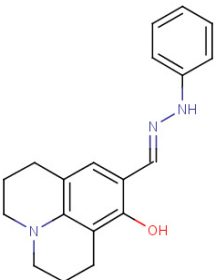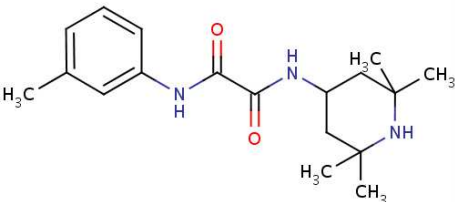
Remaining compounds were investigated for docking to active site and bearing reactive species close to active residues. As a result, molecules on non-steroidal (Table 4.7) and steroidal (Table 4.8) are selected as promising molecules for further detailed docking analysis. Docking positions and interactions of these molecules will be discussed in the next section, along with selection of lead candidates.
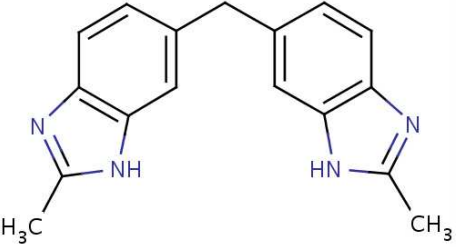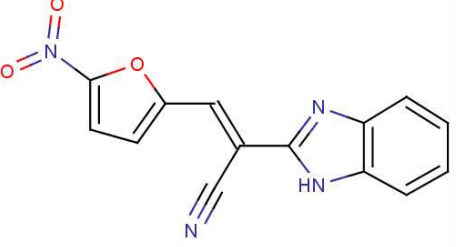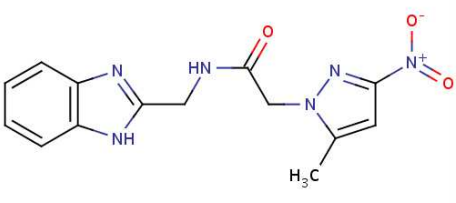
Table 4.7:   Nonsteroidal Candidate Molecules and Energies

ID	Molecule	Docking & Binding Energies
N01		Docking Energy: -11,54 kcal/mol Binding Energy: -10,46 kcal/mol
N02		Docking Energy: -12,47 kcal/mol Binding Energy: -10,93 kcal/mol

N03		Docking Energy: -11,01 kcal/mol Binding Energy: -10,26 kcal/mol
N04		Docking Energy: -10,96 kcal/mol Binding Energy: -10,10 kcal/mol
N05		Docking Energy: -10,72 kcal/mol Binding Energy: -10,02 kcal/mol
N06		Docking Energy: -10,04 kcal/mol Binding Energy: -10,04 kcal/mol

N07		Docking Energy: -11,02 kcal/mol Binding Energy: -10,06 kcal/mol
N08		Docking Energy: -11,63 kcal/mol Binding Energy: -10,81 kcal/mol
N09		Docking Energy: -11,58 kcal/mol Binding Energy: -10,13 kcal/mol
N10		Docking Energy: -11,52 kcal/mol Binding Energy: -10,47 kcal/mol

N11		Docking Energy: -11,35 kcal/mol Binding Energy: -10,14 kcal/mol
N12		Docking Energy: -10,75 kcal/mol Binding Energy: -10,00 kcal/mol
N13		Docking Energy: -10,96 kcal/mol Binding Energy: -10,16 kcal/mol
N14		Docking Energy: -11,88 kcal/mol Binding Energy: -10,07 kcal/mol

N15		Docking Energy: -11,58 kcal/mol Binding Energy: -10,24 kcal/mol
N16		Docking Energy: -10,57 kcal/mol Binding Energy: -10,13 kcal/mol
N17		Docking Energy: -10,17 kcal/mol Binding Energy: -10,17 kcal/mol
N18		Docking Energy: -12,20 kcal/mol Binding Energy: -10,14 kcal/mol

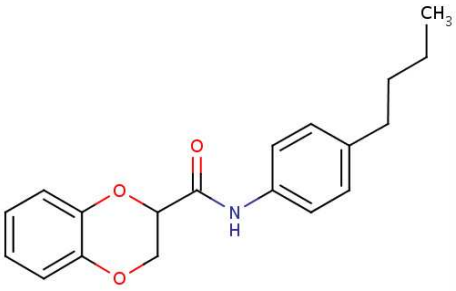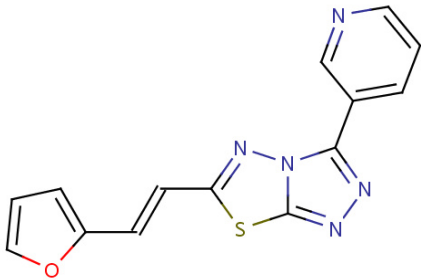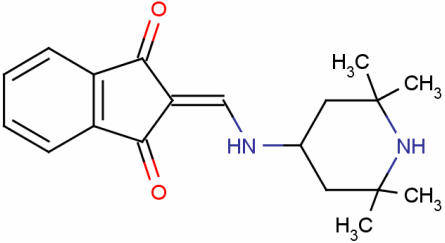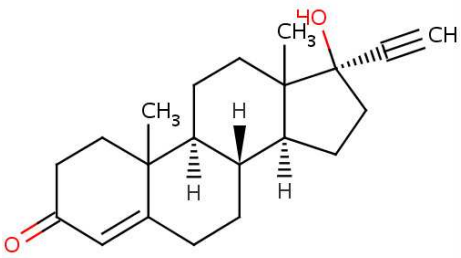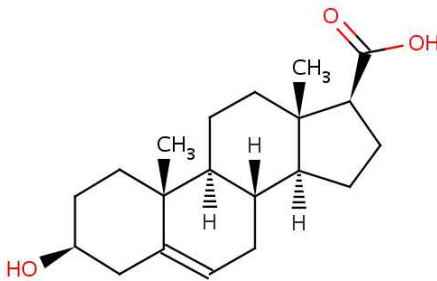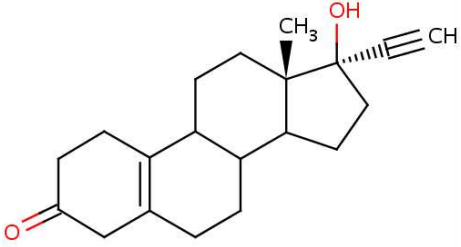| N19		Docking Energy: -11,30 kcal/mol Binding Energy: -10,07 kcal/mol
N20		Docking Energy: -11,30 kcal/mol Binding Energy: -10,07 kcal/mol

Table 4.8:   Steroidal Candidate Molecules and Energies

ID	Molecule	Docking & Binding Energies
S1		Docking Energy: -11,08 kcal/mol Binding Energy: -10,72 kcal/mol
S2		Docking Energy: -10,83 kcal/mol Binding Energy: -10,51 kcal/mol
S3		Docking Energy: -10,51 kcal/mol Binding Energy: -10,14 kcal/mol

4.4.   Detailed Docking and Selection of Candidates

Using filters for molecular weight and virtual screening a library of ~600000 compounds were reduced to 23 molecules. Among these molecules, 20 of them were non-steroidal and 3 of them were steroidal compounds. Results of detailed docking for these molecules are presented on Table 4.9 with comparison of best docking energy values for virtual screening:

Table 4.9:   Results of Virtual Screening and Detailed Docking for Candidates
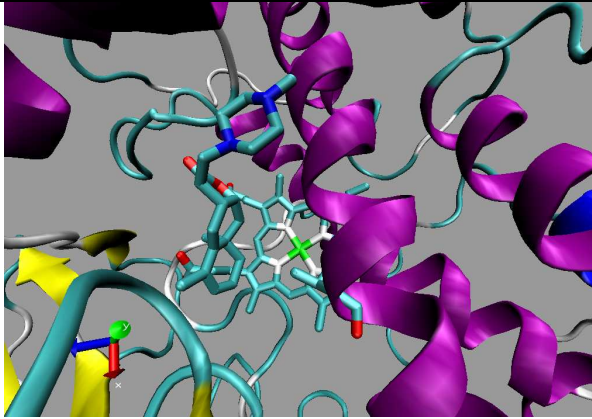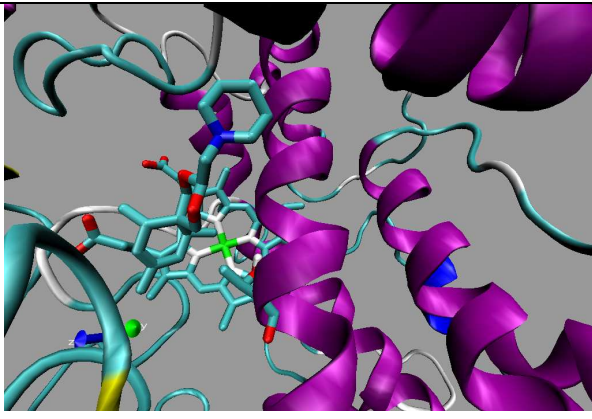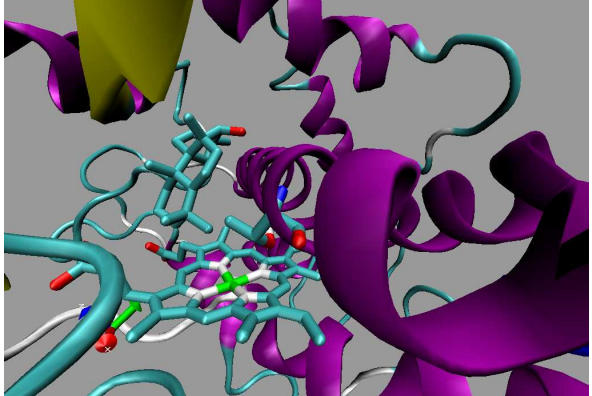
ID	Virtual Screening	Detailed Docking
Nonsteroidal Molecules		
N01	-11,54 kcal/mol	-11,99 kcal/mol
N02	-12,47 kcal/mol	-13,24 kcal/mol
N03	-11,01 kcal/mol	-11,93 kcal/mol
N04	-10,96 kcal/mol	-11,30 kcal/mol
N05	-10,72 kcal/mol	-11,30 kcal/mol
N06	-10,04 kcal/mol	-10,81 kcal/mol
N07	-10,96 kcal/mol	-11,02 kcal/mol
N08	-9,12 kcal/mol	-11,11 kcal/mol
N09	-11,58 kcal/mol	-11,77 kcal/mol
N10	-11,52 kcal/mol	-11,93 kcal/mol
N11	-11,35 kcal/mol	-11,83 kcal/mol
N12	-10,75 kcal/mol	-10,82 kcal/mol
N13	-10,96 kcal/mol	-11,77 kcal/mol

N14	-11,88 kcal/mol	-12,25 kcal/mol
N15	-11,58 kcal/mol	-12,07 kcal/mol
N16	-10,57 kcal/mol	-12,26 kcal/mol
N17	-10,17 kcal/mol	-11,91 kcal/mol
N18	-12,20 kcal/mol	-12,43 kcal/mol
N19	-11,30 kcal/mol	-11,72 kcal/mol
N20	-11,30 kcal/mol	-12,76 kcal/mol
Steroidal Molecules		
S1	-10,52 kcal/mol	-11,91 kcal/mol
S2	-11,08 kcal/mol	-11,72 kcal/mol
S3	-11,10 kcal/mol	-11,63 kcal/mol
S4	-10,83 kcal/mol	-11,36 kcal/mol
S5	-10,03 kcal/mol	-10,21 kcal/mol
S6	-10,09 kcal/mol	-11,12 kcal/mol
S7	-10,51 kcal/mol	-11,14 kcal/mol

As seen on the table, docking scores for all molecules have improved as expected with finer setting of parameters and increased grid resolution. Among nonsteroidal molecules; N3, N4, N6, N13, and N19 do not offer well positioned reactive species for

heme iron or catalytic threonine residue. Docking positions for these molecules are shown on Table 4.10.

Table 4.10:   Docking Conformations for Rejected Nonsteroidal Molecules

N3	
N4	
N6	

| N13 | |
| N19 | |

Rejected nonsteroidal molecules lacked an oxygen or nitrogen atom that can interact with heme iron or oxygen of threonine residue responsible for catalytic activity of the enzyme. Steroid based molecules are not suggested as good candidates for selective inhibition of CYP17. As explained in introduction chapter, many enzymes and receptors are able to bind to steroid molecules. Hence, risk of creating side effects is higher for steroidal molecules. Therefore, only three such molecules were used in further studies. Docking configurations of the candidates are shown on Table 4.11:

Table 4.11:   Docking Configurations for Candidate Molecules

Docking Configurations for Candidate Molecules
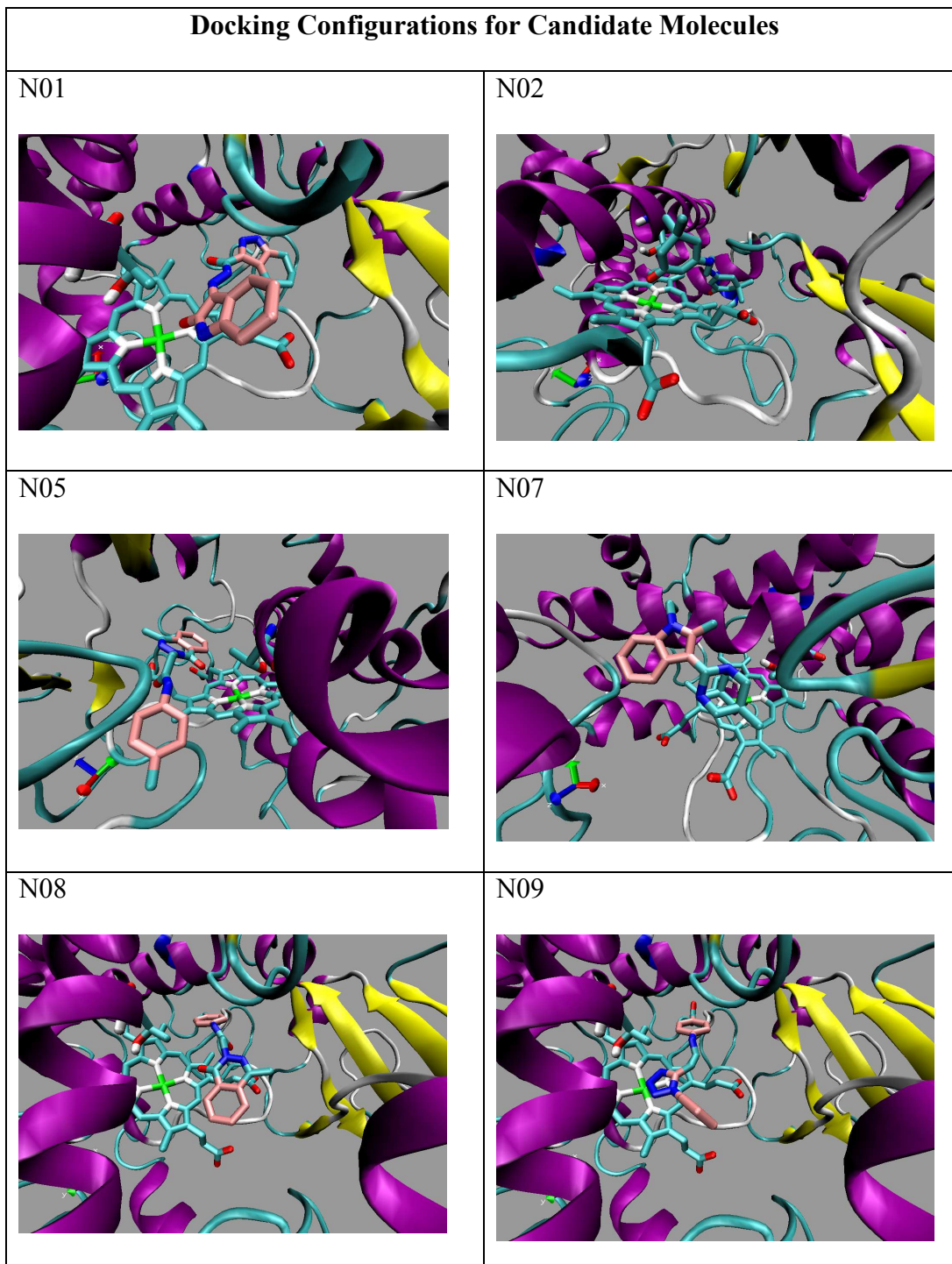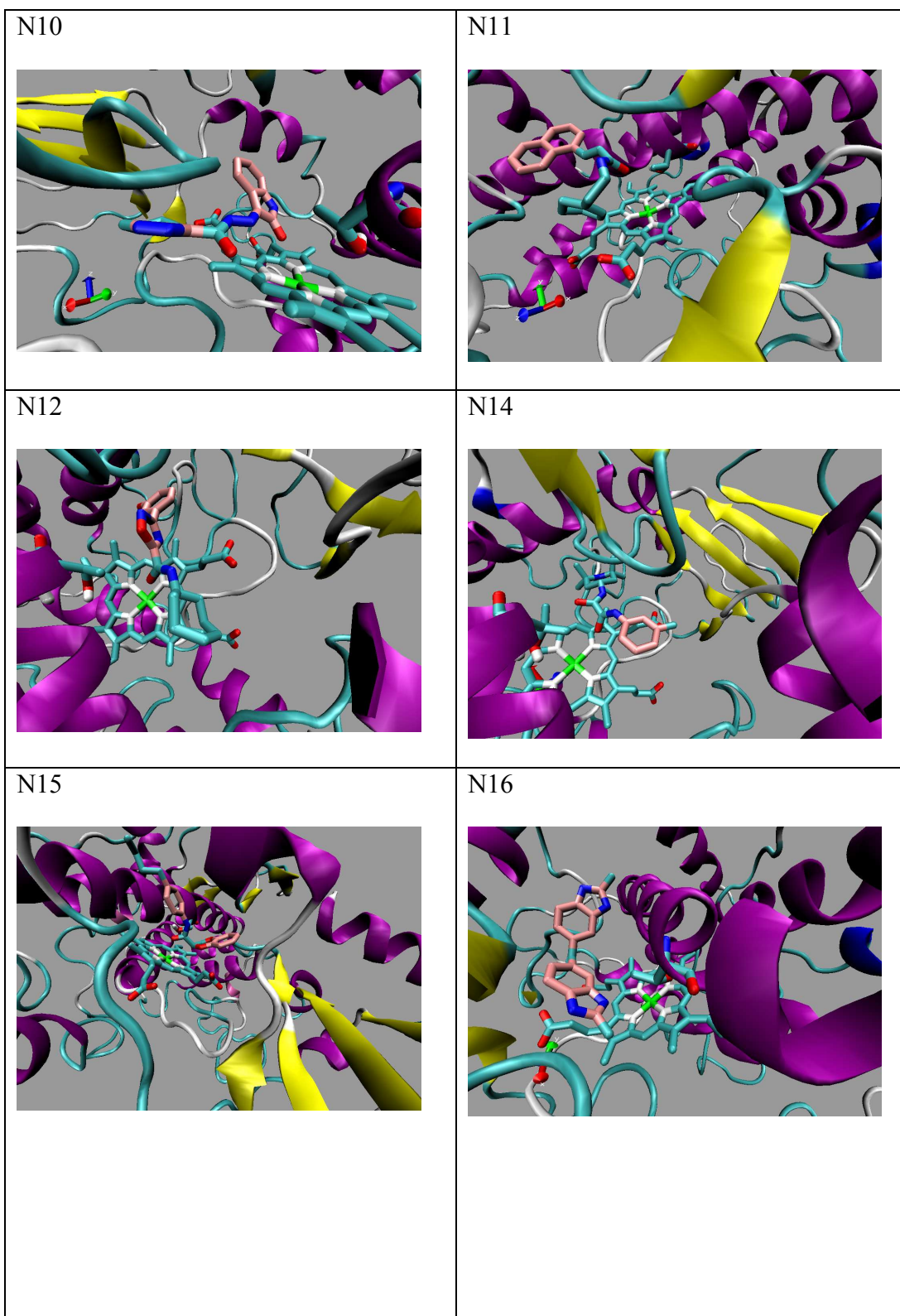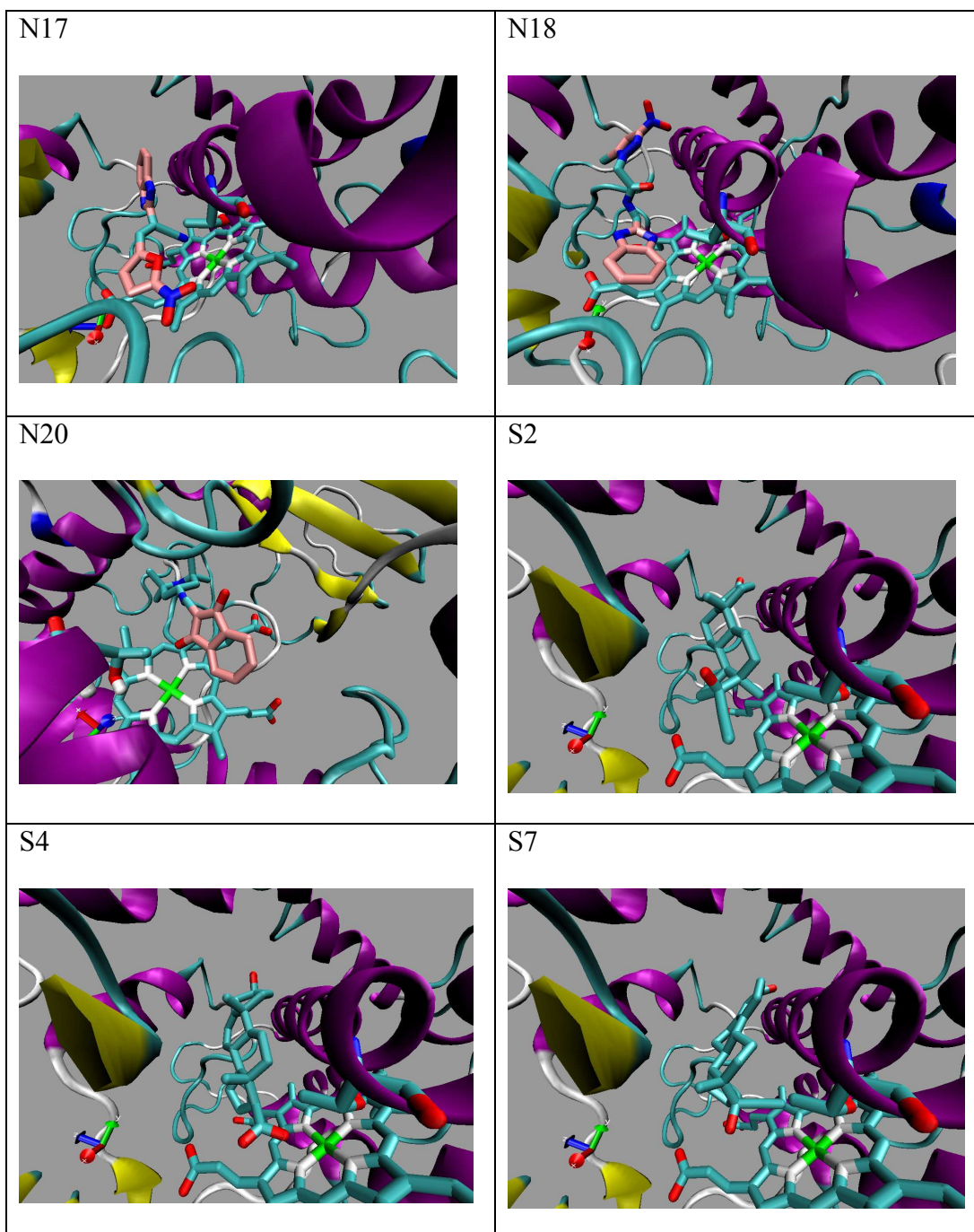
As shown on the table, all nonsteroidal molecules have oxygen or nitrogen atom well oriented with heme iron and threonine oxygen atom, except for N07. This molecule has an unusual ring configuration, forming a structure similar to a cube. When docked to

CYP17 this molecule occupies the active site completely. Therefore, it is selected as a candidate. All selected molecules have comparable energy levels with substrates of CYP17, which increases their chance of competitive inhibition.

4.5.   Expression of human CYP17 in Human Embryonic Kidney Cell Line (HEK-293T)

Different approaches have been taken to measure inhibitory effects of candidate drugs using different enzyme sources. Tissues from mice [49] and rats [50] have been used as a source for the protein as well as human tissues[51]. Although CYP17 protein obtained from animal tissues has significant sequence similarity, they still have differences in activity and substrate selectivity [52]. Therefore, inhibition tests performed using such methods can not be fully translated into effects on human CYP17. On the other hand, enzyme obtained from human tissues is very limited and can not be applicable for routine drug discovery studies. Therefore, isolation of CYP17 gene from human has enabled use of mammalian cell cultures and obtains precise on inhibitory effect of drugs. To this effort first I have cloned human CYP17 cDNA into mammalian expression vector (pCDNA4) using PCR. Amplified full length of PCR products can be seen in Figure 4.3 lane 2 and 4. After sub-cloning into pCDNA4, the success of the sub-cloning was measured by restriction analysis. As shown in Figure 4.3 I have obtained expected fragment from CYP17, 1.2 kb, and 5,6 kb from remaining plasmid. Construct plasmid map is shown in Figure 4.4 and is named as pcDNA4+CYP17.
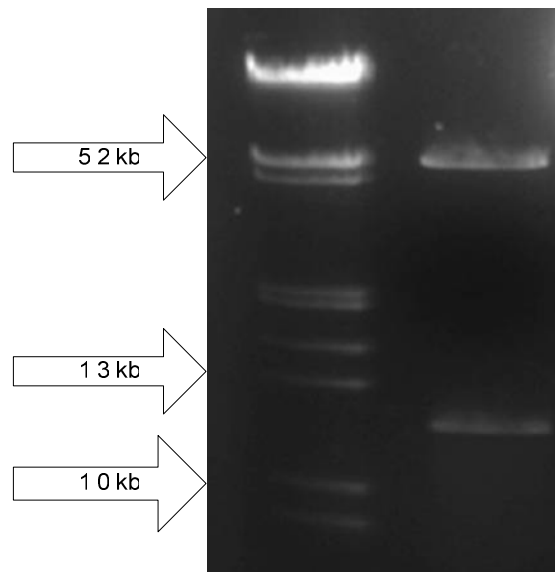
Figure 4.3:   Restriction digestion of pcDNA4+CYP17 with PstI
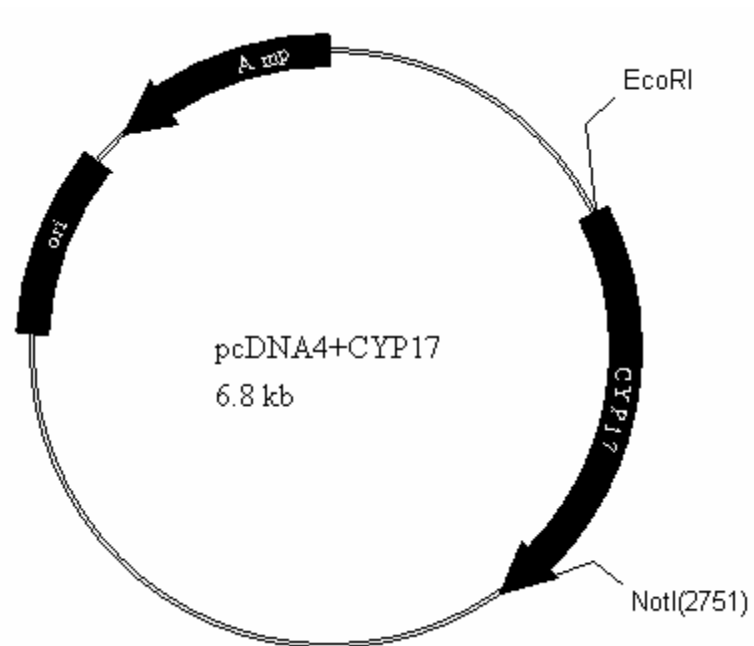
Figure 4.4:   pcDNA4+CYP17 Plasmid Map

To see whether mammalian expression construct express or not in HEK-293T I have transfected pcDNA4+CYP17HEK-293T cells using calcium phosphate transfection and FuGene transfection kit. As shown Figure 4.5, human CYP17 expressed in HEK-293 cell line using anti his antibody:

49 kDa

1st lane
-CYP17
Untransformed

2nd lane
+CYP17
Transformed
by $CaPO_4$
method

3rd lane
+CYP17
Transformed
by FuGENE

1   2   3
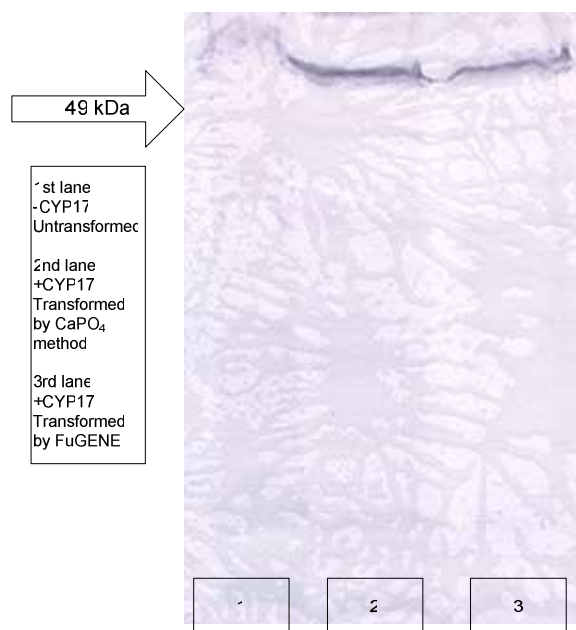
Figure 4.5:   CYP17 Expression with pcDNA4+CYP17 using HEK-293T

The first lane on figure shows an untransformed cell, not able to express CYP17. 2nd lane is a sample from $CaPO_4$ precipitation transfection of pcDNA4+CYP17 to Hek-293T. On this lane expression of CYP17 protein is confirmed with a band with molecular weight over 49kDa. Using FuGENE transfection kit has also showed similar results. This western blot experiment proves that we have successfully cloned and expressed human CYP17 in mammalian cell line and we can proceed for the testing of candidate drugs.

4.6.   Measuring the Inhibitory Effects of Candidate Drugs on Human CYP17

Inhibition of candidate molecules are tested on pcDNA4+CYP17 transformed HEK-293T cell by acetic acid release assay (AARA). Using intact cells in assay has particular

advantages over using purified systems. Since, mammalian cells are used; delivery of the compound will be under question along with inhibition. Therefore, compounds without proper delivery pathways will be eliminated. Moreover, molecules with low stability in cellular conditions will be eliminated as well. Initially, 18 candidate compounds were identified in-silico as potential candidates.

For all 18 molecules, first inhibition test was performed at a concentration of 50 μM. As results preliminary assay, two molecules resulted in inhibition of CYP17 enzyme. These molecules, N15 and S3, are further characterized with inhibition test at various concentrations (Table 4.12, Figure 4.6). One of these molecules were nonsteroidal and the other is a steroid-based compound. For both of these compounds, concentrations covering inhibition pattern are found and listed on table. IC50 value for N15 is lower than 35 μM and for S3, IC50 is lower than 55 μM. There are several compounds published in literature with lower IC50 values. For instance, lowest reported IC50 was 36 nM for a steroido-mimetic compound [11]. However, this compound performed poorly on animal test, by having a very short half-life. The IC50 values for compounds found with this thesis are not drug candidates. These molecules are lead compounds as suggested in structure-based drug design scheme. The expected inhibition level for such lead compounds are at micro-molar concentrations. Further optimizations are the necessary steps for improving inhibition efficiency and reaching to IC50 values at nano-molar concentrations.
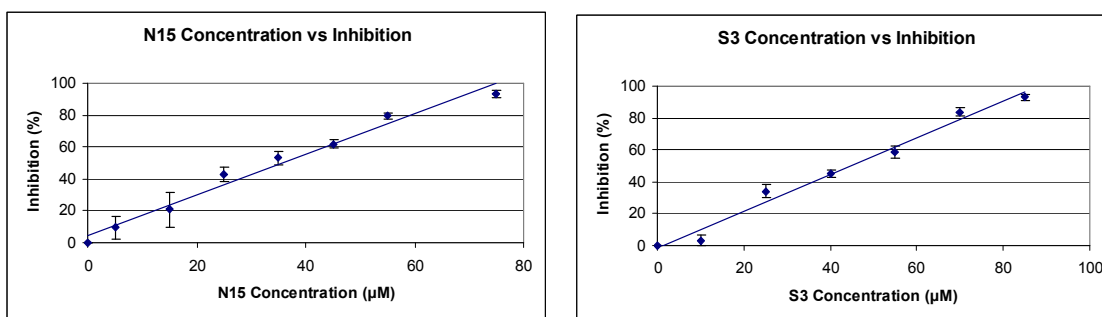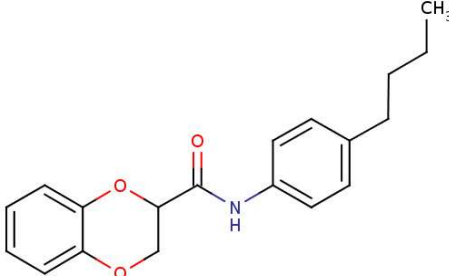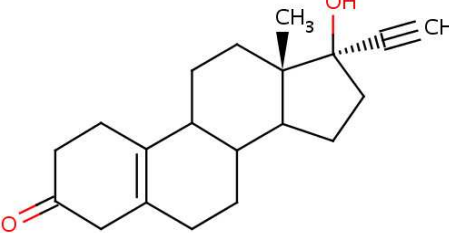
Figure 4.6:   Inhibition of CYP17 Activity by N15 and S3

Table 4.12:   Inhibition of CYP17 Activity by N15 and S3

ID	Drug Concentration (μM)	Average Remaining Activity of CYP17 (%)
N15	0	100
	5	90,6
	15	79,3
	25	57,0
	35	46,8
	45	38,2
	55	20,6
	75	6,6
S3	0	100
	10	96,7
	25	66,1
	40	54,8
	55	41,4
	70	16,4
	85	7,1

4.7.   Cell Viability Assays for Active Candidates

N15 and S3 are actively inhibiting reactions catalyzed by CYP17. Toxicity of a drug candidate is as important as its inhibition potential; therefore, cell viability assays are

performed for N15 and S3. These assays are performed on untransformed HEK-293T cells, and survival of these cells is measured by luminescence. Results are shown with Figure 4.7.

Cell Viability Assay

	50 μM	100 μM	200 μM
N15	70	36,2	35,4
S3	23,4	16,3	12,5

Candidate Concentration (μM)
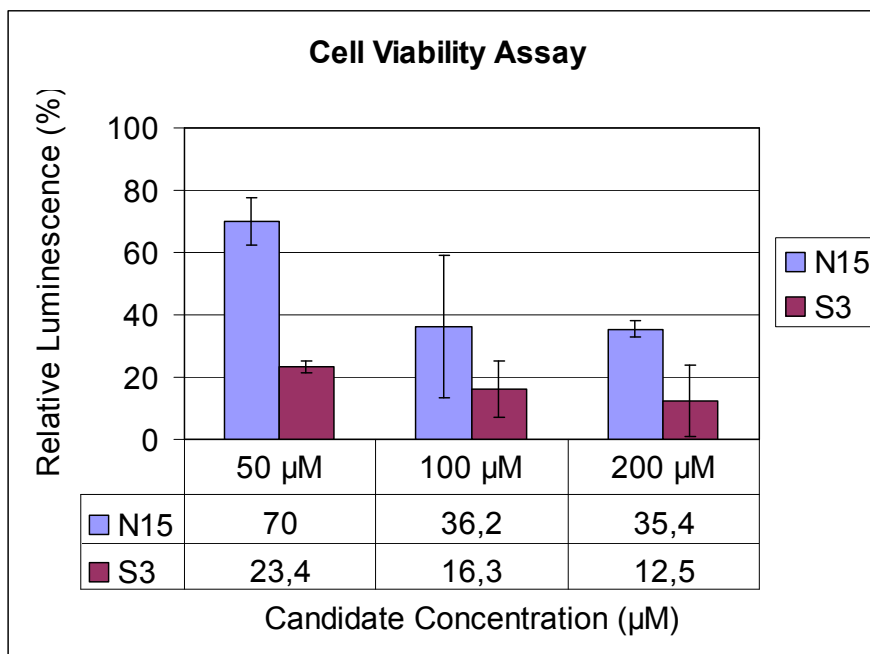
Figure 4.7:   Cell Viability Assay Results

According to results of viability assay, TC50 values are calculated. TC50 represents the concentration of a test compound at which 50% of the cells are killed. Steroid based S3 molecule has very toxic effect on HEK-293T cells even at concentrations close to its IC50 value. The steroid based compound has a TC50 value less than 10 μM. Toxicity value lower than 25 μM for lead compounds is not favored. [53] Therefore, S3 is not a promising candidate for further optimization studies. On the other hand, N15 has fairly good viability scores above its IC50 value. The TC50 for N15 is 88 μM, which is very promising. Therefore, it is used for further optimizations to improve inhibition efficiency and reduce toxicity.

4.8.   Optimization of N15

N15 has promising docking scores on CYP17 active site. Interaction regions for N15 and specific amino acids of CYP17 are shown on Figure 4.8.
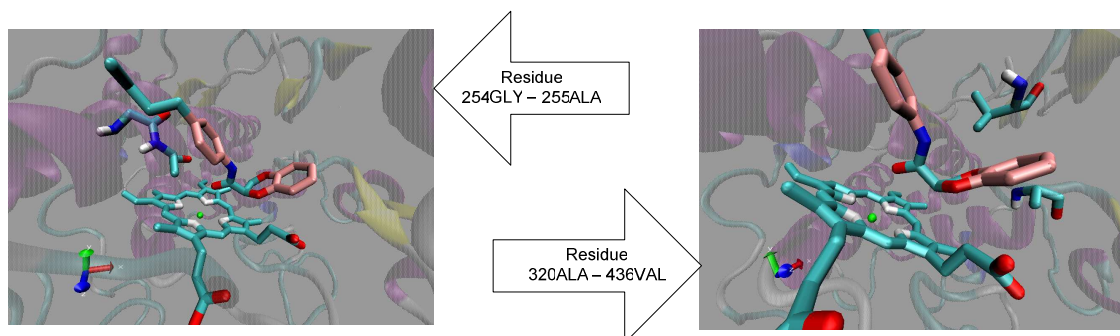
Figure 4.8:   Predicted Interactions of CYP17 and N15

Figure on the left show the interaction of ring and short carbon tail with residues 254 and 255 of I-helix. These residues are glycine and alanine, and their overall hydrophobicity would help stabilize drug molecule in active site. Figure on theright hand show interaction of double ring structure with amino acids 320 and 436. These rings are located between alanine and valine residues, which are hydrophobic. This would also help stabilization of the drug molecule in active site. On both figures, the oxygen atom is well positioned with heme iron, creating posibility of covalent interaction as well.
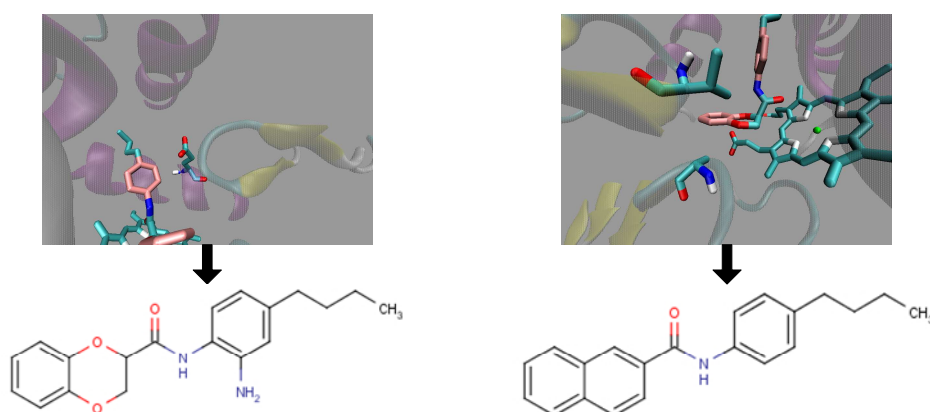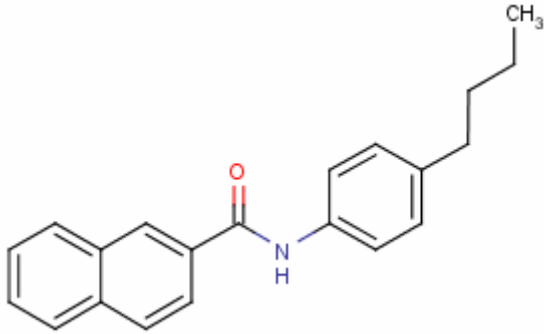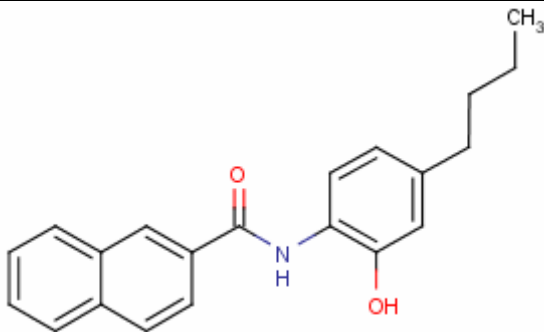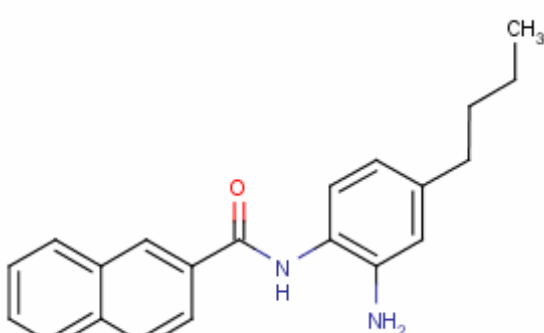
Figure 4.9:   Possible Improvements on N15

On Figure 4.9, two possible improvements on N15 are shown. The left figure shows a possible addition of charged interaction. With the addition of an $-NH_2$ group on single ring, the oxygen atoms of residue 258 (glutamate) may be employed for hydrogen bonding. On the right side, interaction with residues 320 and 436 are view from a different angle. This figure point a possible disturbance due to oxygen atoms on double-ring in between two hydrophobic residues. Possible improvement is shown on the fugure as replacement with carbon atoms and aromatization.

On Table 4.13, three derivative of N15 are shown. The first derivative has replacement of oxygen atoms on double-ring and aromatization, which creates the naphtalene rings. This modifications has improved favored docked structures energy to -0,29 kcal/mol. Second derivative has, also, addition of $-OH$ to single ring. This creates stronger interaction, resulting in improvement of energy by -0,53 kcal/mol. Final derivative has $-NH_2$ in place of -OH, which improves the interaction energy by -0,96 kcal/mol with respect to the original structure.

Table 4.13:   Derivatives of N15 and Improvement in Energy

Derivative Molecule	Improvement in Docking Energy
	-0.29 kcal/mol
	-0,53 kcal/mol
	-0,96 kcal/mol

Chapter 5

Conclusions

Computational and experimental biology methods are used for structure-based drug design for prostate cancer. As a method of treatment, reducing levels of androgens was adopted. In order to achieve this, CYP17 is selected as the target enzyme due to its key role in androgen biosynthesis. Model structure for this enzyme is obtained from protein data bank. Molecular dynamics simulation was used to improve and validate the structure for docking studies. Results of MD simulations revealed that the overall structure is stable, especially in the active site region. It is observed that, relatively less stable regions were located on the surface of the protein on random loops.

MD simulation provided the structure for Autodock docking analysis. Initial dockings were performed on four substrates of the enzyme. Results showed that, all substrates docked successfully to the active site with good orientations with respect to heme and catalytically active threonine residue. Moreover, predicted docking energies indicated strong affinity of these substrates to CYP17.

A library of 600.000 compounds was used to identify new inhibitor candidate molecules. The number of compounds was reduced to ~70.000 by applying a molecular weight filter. Due to relatively small active site architecture, limit was selected as 350 g/mol. This cut-off allows compounds larger than substrates which have molecular weights around 310 g/mol. After virtual screening, molecules giving to docking scores were filtered again based on their binding energy. This filter is necessary to remove long hydrocarbon based molecules. These molecules have high torsional freedom, which results in docking to any site efficiently. For this reason, they can not be specific inhibitors. Their torsional freedom brings instability, which is reflected on increased binding energy. By using this knowledge, molecules with low binding scores are

eliminated regardless of their docking energy. This filter reduced the number of compounds to mere hundreds. Remaining molecules are examined for their binding conformation and reactive species they possess. Several molecules were eliminated due to docking far from active site, as well as many compounds lacking reactive oxygen or nitrogen atom that have potential to interact with heme iron. Finally, a list with 23 molecules, 20 nonsteroidal and 3 steroidal, were obtained. The number of steroidal compounds was kept small due to their disadvantage on side effect risk.

Compounds obtained by coarse docking scores of virtual screening were put in fine docking setup for detailed analysis. As expected, docking scores were improved with higher resolution grids and wider search parameters. However, some molecules still lacked reactive species favorable oriented towards the center of active site. Therefore, list for non steroidal compounds were reduced to 15.

In the experimental section, necessary expression systems for CYP17 enzyme synthesis were done. Mammalian expression vectors were prepared. Vector designed for mammalian cell expression, pcDNA4+CYP17, is capable of inducing expression of CYP17 and full length protein is obtained. This method is used as the source for CYP17 in activity assays.

Selected molecules obtained after detailed dockings were tested for inhibition potential. It is found that one nonsteroidal, N15, and one steroidal, S3, molecule was capable of inhibiting CYP17 enzyme activity. IC50 values, which mean minimum concentration for reducing enzymes activity to 50%, were in μM level. As pointed in overview chapter, these molecules are potential candidates and can be optimized for increased inhibition efficiency. The rest of the molecules failed to inhibit CYP17 activity. One of the reasons for this can be delivery efficiency. Since tests are done directly on HEK-293T cell that can express CYP17, these molecules need to pass through cell membrane. Some of these molecules may not be able to penetrate through this barrier. Another reason can be instability of molecule in assay environment. Some molecules may be destroyed, by simply the aqueous medium or by host cell systems,

before they can exert inhibition effect on CYP17. Finally, these molecules can have stronger affinity towards other proteins, especially other cytochromes. As the template protein for the model of CYP17, cytochromes are a large family of proteins sharing sequence and structure similarity. It is also possible that, these molecules would interact with other cytochromes rather than CYP17.

Compounds with inhibitory effects were tested for their toxic effects. Cell viability assays are performed for N15 and S3. Results show that steroidal molecule, S3, has high toxicity even close to its IC50. Therefore, S3 has not a good potential as a drug. On the other hand, N15 has encouraging results for viability of cells. This nonsteroidal molecule can be used as a drug candidate for optimization. Three possible modifications are listed with their improvements in docking energy for this molecule. These modifications can decrease IC50 values as well as decreasing toxic effects.

In this study, structure-based drug design method is used with a computational model structure of CYP17. A grid-based docking strategy is used for identification and analysis of new drug candidates for prostate cancer. Results shows that two novel inhibitors are discovered by virtual screening and detailed analysis with this strategy. One non-steroidal molecule has low toxicity and potential for optimization in inhibition efficiency, as well as toxicity.

In future studies, new virtual libraries can be scanned with the same virtual screening procedure. For detailed analysis, newer docking algorithms with ligand and receptor flexibility can be used to improve success rate. Molecular dynamics simulations of protein and docked drug molecule can be used to understand interactions in between better. This would also help suggest more accurate modifications on lead molecules to improve drug efficiency.

Supplementary 1

Minimization and Equilibration Configuration File for NAMD2

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

# Minimization of
# CYP17 in a Water Box
# Minimization (w/ fixed atoms), heating (w/ restraint),
# and equilibration (w/ decreasing restraints, and 60000 step
# w/out restraints).

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

structure        ../common/cyp17.psf
Coordinates      ../common/cyp17.pdb
#Velocities      ../mini/equil_ca_0.vel
#extendedSystem     ../mini/equil_ca_0.xsc

set temperature       0
temperature       $temperature
#set outputname      mini
#set restartname  mini
#firsttimestep      0

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

# Input

paraTypeCharmm        on
parameters       ../common/par_all27_prot_lipid_wFE-S.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          0.0000001
nonBondedFreq       1
fullElectFrequency   2
stepsPerCycle      10

# Constant Temperature Control

langevin        on
langevinDamping         5
langevinTemp         310
langevinHydrogen    off

# Periodic Boundary Conditions

cellBasisVector1 92.49  00.00  00.00
cellBasisVector2 00.00  90.65  00.00
cellBasisVector3 00.00  00.00  93.41
cellOrigin      -1.37  -8.89  1.72
margin          2.5
wrapAll             on

# PME (for full-system periodic electrostatics)

PME             on
PMEGridsizeX        96
PMEGridsizeY        96
PMEGridsizeZ        96

# Constant Preassure Control (variable volume)

useGroupPressure     yes# smaller fluctuations
useFlexibleCell         no # allow dimensions to fluctuate independently
useConstantRationo # fix shape in x-y plane
```

```
langevinPiston          on
langevinPistonTarget    1.01325 ;# in bar -> 1 atm
langevinPistonPeriod    100
langevinPistonDecay50
langevinPistonTemp 310

# Output

outputEnergies          500
outputPressure          500
xstFreq            500
DCDFreq            500
restartfreq        500
```

##
EXTRA PARAMETERS
##

Fixed Atoms

```
fixedAtoms         on
fixedAtomsForces   on
fixedAtomsFile         ../common/fix_backbone_heme.pdb
fixedAtomsCol      B
```

Constraints

```
constraints        on
consRef            ../common/restrain_ca_heme.pdb
consKFile      ../common/restrain_ca_heme.pdb
consKCol       B
```

##
EXECUTION SCRIPT
##

```
outputname        cyp17_equil_out

# 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

# heat with CAs and heme restrained
# langevin     on
for { set TEMP 0 } { $TEMP < 310 } { incr TEMP 10 } {
    langevinTemp        $TEMP
    #langevinPistonTemp    $TEMP
    run 5000
}
langevinTemp        310
output       heating

# equilibrate volume with CAs restrained
langevinPiston        on
run          5000
output       equil_ca_1

constraintScaling 0.75
run          5000
output       equil_ca_075

constraintScaling     0.50
run              5000
output              equil_ca_050

constraintScaling     0.25
run              5000
output              equil_ca_025

constraintScaling     0
run              90000
output              equil_ca_0
```

Supplementary 2

Simulation Configuration File for NAMD2

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

# Equilibration of
# CYP17 in a Water Box
# After minimization (w/ fixed atoms), heating (w/ restraint),
# and equilibration (w/ decreasing restraints, and 60000 step
# w/out restraints).


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

structure         ../common/cyp17.psf
Coordinates       ../common/cyp17.pdb
binCoordinates        ../mini/equil_ca_0.coor
binVelocities     ../mini/equil_ca_0.vel
extendedSystem        ../mini/equil_ca_0.xsc

set temperature       310
set outputname        run1
set restartname       run1
firsttimestep     285010


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

# Input

paraTypeCharmm        on
parameters        ../common/par_all27_prot_lipid_wFE-S.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 92.49  00.00   00.00
cellBasisVector2 00.00  90.65   00.00
cellBasisVector3 00.00  00.00   93.41
cellOrigin      -1.37  -8.89   1.72
margin          2.5
wrapAll             on

# PME (for full-system periodic electrostatics)

PME             on
PMEGridsizeX        96
PMEGridsizeY        96
PMEGridsizeZ        96

# Constant Preassure Control (variable volume)

useGroupPressure    yes# smaller fluctuations
useFlexibleCell     no # allow dimensions to fluctuate independently
useConstantRationo # fix shape in x-y plane

langevinPiston          on
```

```
langevinPistonTarget      1.01325 ;# in bar -> 1 atm
langevinPistonPeriod      100
langevinPistonDecay50
langevinPistonTemp 310

# Output

outputname          $outputname
restartname         $restartname
outputEnergies      1000
outputPressure      1000
xstFreq             5000
DCDFreq             5000
restartfreq     5000


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

# Fixed Atoms

fixedAtoms          off
fixedAtomsForces    off
fixedAtomsFile          ../common/fix_backbone_heme.pdb
fixedAtomsCol       B

# Constraints

constraints         off
consRef                 ../common/restrain_ca_heme.pdb
consKFile       ../common/restrain_ca_heme.pdb
consKCol        B


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

run                 5000000
```

Supplementary 3

Virtual Screening Grid Parameter Files

```
receptor 2c17md.pdbqs          # macromolecule
gridfld 2c17md.maps.fld        # grid_data_file
npts 52 52 52                  # num.grid points in xyz
spacing 0.375                  # spacing(A)
gridcenter -0.369 -5.707 -5.889     # xyz-coordinates or auto
types CAONSH                   # atom type names – also for PfFcbI
smooth 0.5                     # store minimum energy w/in rad(A)
map 2c17md.C.map               # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6  # C-C lj
nbp_r_eps  3.75 0.0230026 12  6  # C-N lj
nbp_r_eps  3.60 0.0257202 12  6  # C-O lj
nbp_r_eps  4.00 0.0257202 12  6  # C-S lj
nbp_r_eps  3.00 0.0081378 12  6  # C-H lj
nbp_r_eps  3.75 0.0230026 12  6  # C-X lj
nbp_r_eps  4.00 0.0222750 12  6  # C-M lj
sol_par  12.77 0.6844   # C atomic fragmental volume, solvation parameters
constant  0.000     # C grid map constant energy
map 2c17md.A.map               # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6  # A-C lj
nbp_r_eps  3.75 0.0230026 12  6  # A-N lj
nbp_r_eps  3.60 0.0257202 12  6  # A-O lj
nbp_r_eps  4.00 0.0257202 12  6  # A-S lj
nbp_r_eps  3.00 0.0081378 12  6  # A-H lj
nbp_r_eps  3.00 0.0081378 12  6  # A-X lj
nbp_r_eps  2.65 0.0538015 12  6  # A-M lj
sol_par  10.80 0.1027   # A atomic fragmental volume, solvation parameters
constant  0.000     # A grid map constant energy
map 2c17md.O.map               # atom-specific affinity map
nbp_r_eps  3.60 0.0257202 12  6  # O-C lj
nbp_r_eps  3.35 0.0265667 12  6  # O-N lj
nbp_r_eps  3.20 0.0297000 12  6  # O-O lj
nbp_r_eps  3.60 0.0297000 12  6  # O-S lj
nbp_r_eps  1.90 0.3280000 12 10   # O-H hb
nbp_r_eps  3.35 0.0265667 12  6  # O-X lj
nbp_r_eps  3.60 0.0257202 12  6  # O-M lj
sol_par   0.00 0.0000   # O atomic fragmental volume, solvation parameters
constant  0.236     # O grid map constant energy
map 2c17md.N.map               # atom-specific affinity map
nbp_r_eps  3.75 0.0230026 12  6  # N-C lj
```

```
nbp_r_eps  3.50 0.0237600 12  6   # N-N lj
nbp_r_eps  3.35 0.0265667 12  6   # N-O lj
nbp_r_eps  3.75 0.0265667 12  6   # N-S lj
nbp_r_eps  1.90 0.3280000 12 10   # N-H hb
nbp_r_eps  2.75 0.0084051 12  6   # N-X lj
nbp_r_eps  2.40 0.0555687 12  6   # N-M lj
sol_par   0.00 0.0000   # N atomic fragmental volume, solvation parameters
constant  0.000     # N grid map constant energy
map 2c17md.S.map              # atom-specific affinity map
nbp_r_eps  4.00 0.0257202 12  6   # S-C lj
nbp_r_eps  3.75 0.0265667 12  6   # S-N lj
nbp_r_eps  3.60 0.0297000 12  6   # S-O lj
nbp_r_eps  4.00 0.0297000 12  6   # S-S lj
nbp_r_eps  2.50 0.0656000 12 10   # S-H hb
nbp_r_eps  3.00 0.0093852 12  6   # S-X lj
nbp_r_eps  2.65 0.0621176 12  6   # S-M lj
sol_par  0.000 0.000   #S atomic fragmental volume, solvation parameters
constant  0.000        #S grid map constant energy
map 2c17md.H.map              # atom-specific affinity map
nbp_r_eps  3.00 0.0081378 12  6   # H-C lj
nbp_r_eps  1.90 0.3280000 12 10   # H-N hb
nbp_r_eps  1.90 0.3280000 12 10   # H-O hb
nbp_r_eps  2.50 0.0656000 12 10   # H-S hb
nbp_r_eps  2.00 0.0029700 12  6   # H-H lj
nbp_r_eps  2.00 0.0029700 12  6   # H-X lj
nbp_r_eps  1.65 0.0196465 12  6   # H-M lj
sol_par   0.00 0.0000   # H atomic fragmental volume, solvation parameters
constant  0.118     # H grid map constant energy
elecmap 2c17md.e.map              # electrostatic potential map
dielectric -0.1146              # <0, distance-dep.diel;>0, constant
#
```

Supplementary 4

Detailed Docking Grid Parameter Files

```
receptor 2c17md.pdbqs          # macromolecule
gridfld 2c17md.maps.fld        # grid_data_file
npts 126 126 126               # num.grid points in xyz
spacing 0.150                  # spacing(A)
gridcenter -0.369 -5.707 -5.889     # xyz-coordinates or auto
types CAONSH                   # atom type names – also for PfFcbI
smooth 0.5                     # store minimum energy w/in rad(A)
map 2c17md.C.map               # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6  # C-C lj
nbp_r_eps  3.75 0.0230026 12  6  # C-N lj
nbp_r_eps  3.60 0.0257202 12  6  # C-O lj
nbp_r_eps  4.00 0.0257202 12  6  # C-S lj
nbp_r_eps  3.00 0.0081378 12  6  # C-H lj
nbp_r_eps  3.75 0.0230026 12  6  # C-X lj
nbp_r_eps  4.00 0.0222750 12  6  # C-M lj
sol_par  12.77 0.6844   # C atomic fragmental volume, solvation parameters
constant  0.000     # C grid map constant energy
map 2c17md.A.map               # atom-specific affinity map
nbp_r_eps  4.00 0.0222750 12  6  # A-C lj
nbp_r_eps  3.75 0.0230026 12  6  # A-N lj
nbp_r_eps  3.60 0.0257202 12  6  # A-O lj
nbp_r_eps  4.00 0.0257202 12  6  # A-S lj
nbp_r_eps  3.00 0.0081378 12  6  # A-H lj
nbp_r_eps  3.00 0.0081378 12  6  # A-X lj
nbp_r_eps  2.65 0.0538015 12  6  # A-M lj
sol_par  10.80 0.1027   # A atomic fragmental volume, solvation parameters
constant  0.000     # A grid map constant energy
map 2c17md.O.map               # atom-specific affinity map
nbp_r_eps  3.60 0.0257202 12  6  # O-C lj
nbp_r_eps  3.35 0.0265667 12  6  # O-N lj
nbp_r_eps  3.20 0.0297000 12  6  # O-O lj
nbp_r_eps  3.60 0.0297000 12  6  # O-S lj
nbp_r_eps  1.90 0.3280000 12 10   # O-H hb
nbp_r_eps  3.35 0.0265667 12  6  # O-X lj
nbp_r_eps  3.60 0.0257202 12  6  # O-M lj
sol_par   0.00 0.0000   # O atomic fragmental volume, solvation parameters
constant  0.236     # O grid map constant energy
map 2c17md.N.map               # atom-specific affinity map
nbp_r_eps  3.75 0.0230026 12  6  # N-C lj
```

nbp_r_eps 3.50 0.0237600 12 6   # N-N lj
nbp_r_eps 3.35 0.0265667 12 6   # N-O lj
nbp_r_eps 3.75 0.0265667 12 6   # N-S lj
nbp_r_eps 1.90 0.3280000 12 10   # N-H hb
nbp_r_eps 2.75 0.0084051 12 6   # N-X lj
nbp_r_eps 2.40 0.0555687 12 6   # N-M lj
sol_par   0.00 0.0000   # N atomic fragmental volume, solvation parameters
constant 0.000   # N grid map constant energy
map 2c17md.S.map   # atom-specific affinity map
nbp_r_eps 4.00 0.0257202 12 6   # S-C lj
nbp_r_eps 3.75 0.0265667 12 6   # S-N lj
nbp_r_eps 3.60 0.0297000 12 6   # S-O lj
nbp_r_eps 4.00 0.0297000 12 6   # S-S lj
nbp_r_eps 2.50 0.0656000 12 10   # S-H hb
nbp_r_eps 3.00 0.0093852 12 6   # S-X lj
nbp_r_eps 2.65 0.0621176 12 6   # S-M lj
sol_par 0.000 0.000   #S atomic fragmental volume, solvation parameters
constant 0.000   #S grid map constant energy
map 2c17md.H.map   # atom-specific affinity map
nbp_r_eps 3.00 0.0081378 12 6   # H-C lj
nbp_r_eps 1.90 0.3280000 12 10   # H-N hb
nbp_r_eps 1.90 0.3280000 12 10   # H-O hb
nbp_r_eps 2.50 0.0656000 12 10   # H-S hb
nbp_r_eps 2.00 0.0029700 12 6   # H-H lj
nbp_r_eps 2.00 0.0029700 12 6   # H-X lj
nbp_r_eps 1.65 0.0196465 12 6   # H-M lj
sol_par   0.00 0.0000   # H atomic fragmental volume, solvation parameters
constant 0.118   # H grid map constant energy
elecmap 2c17md.e.map   # electrostatic potential map
dielectric -0.1146   # <0, distance-dep.diel;>0, constant
#

Supplementary 5

Virtual Screening Docking Parameter Files

```
outlev 1                    # diagnostic output level
seed pid time                 # seeds for random generator
types CANO                    # atom type names
fld 2c17md.maps.fld            # grid_data_file
map 2c17md.C.map                # atom-specific affinity map
map 2c17md.A.map                # atom-specific affinity map
map 2c17md.N.map                # atom-specific affinity map
map 2c17md.O.map                # atom-specific affinity map
map 2c17md.e.map                 # electrostatics map
move molecule.pdbq                # small molecule
about 0.0 0.0 -0.0001            # small molecule center
tran0 random                  # initial coordinates/A or random
quat0 random                  # initial quaternion
ndihe 4                     # number of active torsions
dihe0 random                  # initial dihedrals (relative) or random
tstep 2.0                   # translation step/A
qstep 50.0                   # quaternion step/deg
dstep 50.0                   # torsion step/deg
torsdof 5 0.3113               # torsional degrees of freedom and coeffiecent
intnbp_r_eps  4.00 0.0222750 12 6    # C-C lj
intnbp_r_eps  4.00 0.0222750 12 6    # C-A lj
intnbp_r_eps  3.75 0.0230026 12 6    # C-N lj
intnbp_r_eps  3.60 0.0257202 12 6    # C-O lj
intnbp_r_eps  4.00 0.0222750 12 6    # A-A lj
intnbp_r_eps  3.75 0.0230026 12 6    # A-N lj
intnbp_r_eps  3.60 0.0257202 12 6    # A-O lj
intnbp_r_eps  3.50 0.0237600 12 6    # N-N lj
intnbp_r_eps  3.35 0.0265667 12 6    # N-O lj
intnbp_r_eps  3.20 0.0297000 12 6    # O-O lj
#
rmstol 2.0                   # cluster_tolerance/A
extnrg 1000.0                  # external grid energy
e0max 0.0 10000                 # max initial energy; max number of retries
ga_pop_size 50                 # number of individuals in population
ga_num_evals 1000000              # maximum number of energy evaluations
ga_num_generations 27000           # maximum number of generations
ga_elitism 1                  # number of top individuals to survive to next generation
ga_mutation_rate 0.02             # rate of gene mutation
ga_crossover_rate 0.8             # rate of crossover
```

```
ga_window_size 10           #
ga_cauchy_alpha 0.0          # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0            # Beta parameter Cauchy distribution
set_ga                # set the above parameters for GA or LGA
sw_max_its 300             # iterations of Solis & Wets local search
sw_max_succ 4              # consecutive successes before changing rho
sw_max_fail 4             # consecutive failures before changing rho
sw_rho 1.0              # size of local search space to sample
sw_lb_rho 0.01             # lower bound on rho
ls_search_freq 0.06           # probability of performing local search on individual
set_sw1              # set the above Solis & Wets parameters
ga_run 10              # do this many hybrid GA-LS runs
analysis              # perform a ranked cluster analysis
```

Supplementary 6

Detailed Docking Parameter Files

```
outlev 1                    # diagnostic output level
seed pid time                 # seeds for random generator
types CANO                    # atom type names
fld 2c17md.maps.fld              # grid_data_file
map 2c17md.C.map                  # atom-specific affinity map
map 2c17md.A.map                  # atom-specific affinity map
map 2c17md.N.map                  # atom-specific affinity map
map 2c17md.O.map                  # atom-specific affinity map
map 2c17md.e.map                # electrostatics map
move molecule.pdbq                # small molecule
about 0.0 0.0 -0.0001            # small molecule center
tran0 random                # initial coordinates/A or random
quat0 random                # initial quaternion
ndihe 4                # number of active torsions
dihe0 random                # initial dihedrals (relative) or random
tstep 2.0                # translation step/A
qstep 50.0                  # quaternion step/deg
dstep 50.0                  # torsion step/deg
torsdof 5 0.3113                # torsional degrees of freedom and coeffiecent
intnbp_r_eps  4.00 0.0222750 12 6   # C-C lj
intnbp_r_eps  4.00 0.0222750 12 6   # C-A lj
intnbp_r_eps  3.75 0.0230026 12 6   # C-N lj
intnbp_r_eps  3.60 0.0257202 12 6   # C-O lj
intnbp_r_eps  4.00 0.0222750 12 6   # A-A lj
intnbp_r_eps  3.75 0.0230026 12 6   # A-N lj
intnbp_r_eps  3.60 0.0257202 12 6   # A-O lj
intnbp_r_eps  3.50 0.0237600 12 6   # N-N lj
intnbp_r_eps  3.35 0.0265667 12 6   # N-O lj
intnbp_r_eps  3.20 0.0297000 12 6   # O-O lj
#
rmstol 2.0                # cluster_tolerance/A
extnrg 1000.0                 # external grid energy
e0max 0.0 10000                 # max initial energy; max number of retries
ga_pop_size 250                 # number of individuals in population
ga_num_evals 15000000              # maximum number of energy evaluations
ga_num_generations 50000            # maximum number of generations
ga_elitism 1                # number of top individuals to survive to next generation
ga_mutation_rate 0.02              # rate of gene mutation
ga_crossover_rate 0.8              # rate of crossover
```

```
ga_window_size 10             #
ga_cauchy_alpha 0.0            # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0             # Beta parameter Cauchy distribution
set_ga                # set the above parameters for GA or LGA
sw_max_its 300             # iterations of Solis & Wets local search
sw_max_succ 4              # consecutive successes before changing rho
sw_max_fail 4             # consecutive failures before changing rho
sw_rho 1.0             # size of local search space to sample
sw_lb_rho 0.01             # lower bound on rho
ls_search_freq 0.06            # probability of performing local search on individual
set_sw1               # set the above Solis & Wets parameters
ga_run 10              # do this many hybrid GA-LS runs
analysis              # perform a ranked cluster analysis
```

Supplementary 7

PCR Reactions

For 50 µl reaction mixture:

Reagents	Quantity
10X Taq Buffer (Fermentas)	5 µl
10X MgCl$_2$ (25 mM)	5 µl
4 mM dNTP Mix	3 µl
Forward Primer (20pmol/ml)	3 µl
Reverse Primer (20pmol/ml)	3 µl
Template DNA (10pg to 1 µg)	~1 µl
Taq DNA Polymerase (1 to 2 units)	~0.1 µl
Sterile & deionized water	Complete to 50 µl

PCR Program:

Temperature	Time	Cycle
95 °C	4 min	1
95 °C	30 sec	
55 °C	30 sec	30
72 °C	2 min	
72 °C	10 min	1
4 °C	∞	1

Supplementary 8

Restriction Digestion and Alkaline Phosphatase Treatment

Following buffers are used for respective enzyme digestions:

Restriction Enzyme	Fermentas Buffer
NotI	Buffer O
PstI	Buffer O
EcoRI	EcoRI buffer
HindIII	Buffer R
BamHI	BamHI buffer
XhoI	Buffer R

Restrictions are carried out at 37 °C water-bath, for 1 hour or more depending on needs.

Alkaline phosphatase treatments are carried out after digestions when necessary. After digestion, 1 µl of Shrimp Alkaline Phosphatase (SAP, Fermentas) is added to mixture and incubated in 37 °C for 30 minutes. SAP is inactivated by incubation at 65 °C for 10 minutes.

Supplementary 9

Ligation with T4 ligase and Heat-Shock Transformation

Fermentas T4 ligase enzymes and buffer are used for ligation. Reactants are as follows:

Reactants	Quantity
Linearized Vector DNA	~10 ng
Insert DNA	~30 ng to 60 ng
10X Ligation Buffer	1 μl
T4 DNA ligase	1 μl
Sterile and deionized water	Complete to 10 μl

Mixture is incubated at room temperature for 90 minutes.

After ligation is completed, DNA is recovered by ethanol precipitation as follows:

1. Add 25 μl 95% EtOH. (Add 2 μl glycogen, optional)

2. Vortex, then incubate at -20 °C for 10 min.

3. Centrifuge on bench top centrifuge at 14K g for 15 min.

4. Wash pellet with 100 μl 70% EtOH by inversion.

5. Centrifuge for 1 min.

6. Dry pellet under hood and dissolve in 5 μl dH$_2$O.

After DNA is recovered, it is transferred to *E. coli* by heat-shock as follows:

1. Thaw competent cells on ice. Keep DNA on ice until cells are thawed.

2. Add DNA to cells under sterile conditions and mix by inversion. Keep cells on ice for 10 min.

3. Transfer cells to water-bath at 42 °C for 90 sec. Then, transfer cell on ice and incubate for 5 min. Centrifuge for 30 sec and dissolve pellet in 1ml LB.

4. Incubate at 37 °C, 200 RPM shaker for 1 hour. Spread cells on selective agar plates and keep in 37 °C incubator overnight.

Supplementary 10

Transfection of HEK-293T Cells by CaPO$_4$ Precipitation

1. Obtain cells at 70% to 80% confluency.

2. Prepare 2 sets of 15 ml falcon tubes:

    a. In one add 0.5 ml 2x HEBS

    b. In other one add 1-20 μg DNA in 450 μl TE. Also add 50 μl of CaCl$_2$ in 10 mM HEPES, pH 7,2.

3. After gentle mixing of DNA drawing solution up and down in pipette, add all to HEBS solution drop-wise. Do these while blowing bubbles into the solution.

4. Incubate mixture at room temperature for 30 min.

5. Add mixture to plate drop-wise, cover all the surface.

6. Incubate in 37 °C, with 5% CO$_2$ and humidity, overnight.

7. Refresh medium.

8. Next day, collect cells.

Bibliography

[1] Frydenberg, M. and S. Wijesinha, *Diagnosing prostate cancer - what GPs need to know.* Aust Fam Physician, 2007. **36**(5): p. 345-7.

[2] Mettlin, C., *Recent developments in the epidemiology of prostate cancer.* Eur J Cancer, 1997. **33**(3): p. 340-7.

[3] Jemal, A., et al., *Cancer statistics, 2007.* CA Cancer J Clin, 2007. **57**(1): p. 43-66.

[4] Latil, A.G., et al., *Prostate carcinoma risk and allelic variants of genes involved in androgen biosynthesis and metabolism pathways.* Cancer, 2001. **92**(5): p. 1130-7.

[5] Ross, R.K., et al., *5-alpha-reductase activity and risk of prostate cancer among Japanese and US white and black males.* Lancet, 1992. **339**(8798): p. 887-9.

[6] Dong, J.T., *Prevalent mutations in prostate cancer.* J Cell Biochem, 2006. **97**(3): p. 433-47.

[7] Steinberg, G.D., et al., *Family history and the risk of prostate cancer.* Prostate, 1990. **17**(4): p. 337-47.

[8] Thompson, I.M. and D.P. Ankerst, *Prostate-specific antigen in the early detection of prostate cancer.* CMAJ, 2007. **176**(13): p. 1853-8.

[9] Simard, J., et al., *Prostate cancer susceptibility genes: lessons learned and challenges posed.* Endocr Relat Cancer, 2003. **10**(2): p. 225-59.

[10] Douglas, J.A., et al., *Identifying susceptibility genes for prostate cancer--a family-based association study of polymorphisms in CYP17, CYP19, CYP11A1, and LH-beta.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(8): p. 2035-9.

[11] Hartmann, R.W., et al., *Inhibition of CYP 17, a new strategy for the treatment of prostate cancer.* Arch Pharm (Weinheim), 2002. **335**(4): p. 119-28.

[12] Metin, T., *Structural Synthesis of Small Molecule Drug Candidates.* FOCAPD 2004, 2004: p. 395-398.

[13] Hardy, L.W. and A. Malikayil, *The impact of structure-guided drug design on clinical agents.* Curr Drug Discov, 2003(Dec): p. 15-20.

[14] Klebe, G., *Recent developments in structure-based drug design.* J Mol Med, 2000. **78**(5): p. 269-81.

[15] Anderson, A.C., *The process of structure-based drug design.* Chem Biol, 2003. **10**(9): p. 787-97.

[16] Carter, B.S., H.B. Carter, and J.T. Isaacs, *Epidemiologic evidence regarding predisposing factors to prostate cancer.* Prostate, 1990. **16**(3): p. 187-97.

[17] McConnell, J.D., *Physiologic basis of endocrine therapy for prostatic cancer.* Urol Clin North Am, 1991. **18**(1): p. 1-13.

[18] Miller, W.L., *Molecular Biology of steroid hormone synthesis.* Endocr Rev, 1998. **9**: p. 295-318.

[19] Cicek, M.S., et al., *Association of prostate cancer risk and aggressiveness to androgen pathway genes: SRD5A2, CYP17, and the AR.* Prostate, 2004. **59**(1): p. 69-76.

[20] Chung, B.C., et al., *Cytochrome P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues.* Proc Natl Acad Sci U S A, 1987. **84**(2): p. 407-11.

[21] Easton, D.F., et al., *Where are the prostate cancer genes?--A summary of eight genome wide searches.* Prostate, 2003. **57**(4): p. 261-9.

[22] Nason, T.F., X.G. Han, and P.F. Hall, *Cyclic AMP regulates expression of the rat gene for steroid 17 alpha-hydroxylase/C17-20 lyase P-450 (CYP17) in rat Leydig cells.* Biochim Biophys Acta, 1992. **1171**(1): p. 73-80.

[23] Nishihara, M., et al., *Hormonal regulation of rat Leydig cell cytochrome P-45017 alpha mRNA levels and characterization of a partial length rat P-45017 alpha cDNA.* Biochem Biophys Res Commun, 1988. **154**(1): p. 151-8.

[24] Orava, M., R. Voutilainen, and R. Vihko, *Interferon-gamma inhibits steroidogenesis and accumulation of mRNA of the steroidogenic enzymes P450scc and P450c17 in cultured porcine Leydig cells.* Mol Endocrinol, 1989. **3**(6): p. 887-94.

[25] Rainey, W.E., et al., *Transforming growth factor-beta inhibits steroid 17 alpha-hydroxylase cytochrome P-450 expression in ovine adrenocortical cells.* Endocrinology, 1990. **127**(4): p. 1910-5.

[26] Auchus, R.J., T.C. Lee, and W.L. Miller, *Cytochrome b5 augments the 17,20-lyase activity of human P450c17 without direct electron transfer.* J Biol Chem, 1998. **273**(6): p. 3158-65.

[27] Laughton, C.A., et al., *A molecular model for the enzyme cytochrome P450(17 alpha), a major target for the chemotherapy of prostatic cancer.* Biochem Biophys Res Commun, 1990. **171**(3): p. 1160-7.

[28] Lin, D., et al., *Modeling and mutagenesis of the active site of human P450c17.* Mol Endocrinol, 1994. **8**(3): p. 392-402.

[29] Auchus, R.J. and W.L. Miller, *Molecular modeling of human P450c17 (17alpha-hydroxylase/17,20-lyase): insights into reaction mechanisms and effects of mutations.* Mol Endocrinol, 1999. **13**(7): p. 1169-82.

[30] 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.

[31] Ideyama, Y., et al., *YM116, 2-(1H-imidazol-4-ylmethyl)-9H-carbazole, decreases adrenal androgen synthesis by inhibiting C17-20 lyase activity in NCI-H295 human adrenocortical carcinoma cells.* Jpn J Pharmacol, 1999. **79**(2): p. 213-20.

[32] Nnane, I.P., et al., *Inhibition of androgen synthesis in human testicular and prostatic microsomes and in male rats by novel steroidal compounds.* Endocrinology, 1999. **140**(6): p. 2891-7.

[33] Arlt, W., R.J. Auchus, and W.L. Miller, *Thiazolidinediones but not metformin directly inhibit the steroidogenic enzymes P450c17 and 3beta -hydroxysteroid dehydrogenase.* J Biol Chem, 2001. **276**(20): p. 16767-71.

[34] Arlt, W., et al., *Cinnamic acid based thiazolidinediones inhibit human P450c17 and 3beta-hydroxysteroid dehydrogenase and improve insulin sensitivity independent of PPARgamma agonist activity.* J Mol Endocrinol, 2004. **32**(2): p. 425-36.

[35] 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.

[36] Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-52.

[37] Vitkup, D., et al., *Solvent mobility and the protein 'glass' transition.* Nat Struct Biol, 2000. **7**(1): p. 34-8.

[38] Dobson, C.M., A. Sali, and M. Karplus, *Protein folding: a perspective from theory and experiment.* Angew. Chem. Int. Ed. Engl., 1998. **37**: p. 868-893.

[39] Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

[40] MacKerell, J., A. D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, , 1998. **102**: p. 3586-3616.

[41] Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

[42] Collins, J.R., D.L. Camper, and G.H. Loew, *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.

[43] Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD.* Proteins, 2003. **52**(4): p. 609-23.

[44] Meng, E.C., et al., *Orientational sampling and rigid-body minimization in molecular docking.* Proteins, 1993. **17**(3): p. 266-78.

[45] McGann, M.R., et al., *Gaussian docking functions.* Biopolymers, 2003. **68**(1): p. 76-90.

[46] Schellhammer, I. and M. Rarey, *FlexX-Scan: fast, structure-based virtual screening.* Proteins, 2004. **57**(3): p. 504-17.

[47] *MOE 2002.03.* 2002, The Chemical Computing Group Inc. Montreal.

[48] Garrett M. Morris, D.S.G.R.S.H.R.H.W.E.H.R.K.B.A.J.O., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.

[49] Tagawa, N., et al., *17alpha-hydroxylase/C17-20 lyase cytochrome P450 mRNA expressions and enzyme activities during the development of arthritis in collagen-induced arthritis mice.* Biol Pharm Bull, 2004. **27**(10): p. 1663-5.

[50] Kibaly, C., C. Patte-Mensah, and A.G. Mensah-Nyagan, *Molecular and neurochemical evidence for the biosynthesis of dehydroepiandrosterone in the adult rat spinal cord.* J Neurochem, 2005. **93**(5): p. 1220-30.

[51] Klus, G.T., et al., *Growth inhibition of human prostate cells in vitro by novel inhibitors of androgen synthesis.* Cancer Res, 1996. **56**(21): p. 4956-64.

[52] Imai, T., et al., *Expression and purification of functional human 17 alpha-hydroxylase/17,20-lyase (P450c17) in Escherichia coli. Use of this system for study of a novel form of combined 17 alpha-hydroxylase/17,20-lyase deficiency.* J Biol Chem, 1993. **268**(26): p. 19681-9.

[53] Benbow, J.W., et al., *Synthesis and Evaluation of Dinitroanilines for Treatment of Cryptosporidiosis.* Antimicrobial Agents and Chemotherapy, 1998. **42**(2): p. 339-43

VITA

Muhittin Emre Özdemir was born in Konya, Turkey on October 20, 1982. He is an alumnus of Özel Atafen Lisesi, Kocaeli. He received his Bachelor of Science Degree in Molecular Biology and Genetics from Middle East Technical University, Ankara, in June 2005.

From 2005 to 2008, he was a research and teaching assistant in the Computational Sciences and Engineering Department of Koç University. His research includes drug design studies for prostate cancer.