8-2020

Unbinding of Testosterone from the Androgen Receptor and Two BF-3 Site Mutants

Muniruzzaman Chowdhury
The University of Texas Rio Grande Valley

UNBINDING OF TESTOSTERONE FROM THE ANDROGEN RECEPTOR AND TWO BF-3 SITE MUTANTS

A Thesis

by

MUNIRUZZAMAN CHOWDHURY

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2020

Major Subject: Chemistry

UNBINDING OF TESTOSTERONE FROM THE ANDROGEN RECEPTOR AND TWO BF-3 SITE MUTANTS

A Thesis
by
MUNIRUZZAMAN CHOWDHURY

COMMITTEE MEMBERS

Dr. Evangelia Kotsikorou
Chair of Committee

Dr. Frank Dean
Committee Member

Dr. Jason Parsons
Committee Member

Dr. Javier Gutierrez
Committee Member

August 2020

ABSTRACT

Chowdhury, Muniruzzaman, Unbinding of Testosterone from the Androgen Receptor and Two BF-3 Site Mutants. Master of Science (MS), August, 2020, 41 pp, 5 tables, 19 figures, 32 references, 22 titles.

The androgen receptor (AR) can be activated by molecules binding in the steroid binding pocket. It can also be regulated allosterically by cofactors binding to the AF-2 surface site as well as by exogenous small molecules binding to the BF-3 surface site. Recent data indicated that mutations in the BF-3 site changed the amount of the endogenous steroid needed to activate the receptor, indicating that mutations of BF-3 can cause an allosteric effect. Molecular dynamics and steered molecular dynamics simulations were used to study the allosteric effect of the mutations on AR structure and the unbinding pathways of testosterone (TES) from AR. It was found that the BF-3 mutations did not have the destabilizing effect expected. However, the mutations resulted in changes of the hydrogen bonding patterns for TES bound in the steroid binding pocket as well as differences in the unbinding pathways for TES.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The androgen receptor (AR), a member of the nuclear receptor superfamily, belongs to the steroid receptor subfamily and its endogenous ligands are testosterone (TES) and dihydrotestosterone (DHT). The superfamily of Nuclear receptors (NRs) also includes the glucocorticoid receptor (GR/NR3C1), the mineralocorticoid receptor (MR/NR3C2), the progesterone receptor (PR/NR3C3) and the estrogen receptors α and β (ERα/NR3A1; ERβ/NR3A2). Steroid receptors are very popular therapeutic targets because they play a pivotal role in a number of endocrine-related diseases [1,2]. AR is activated by the binding of TES or DHT in the cytoplasm and then it's translocated into the nucleus [3] .

The primary function of the androgen receptor is that it's a DNA-binding transcription factor that is responsible for the regulation of gene expression; [4] however, it has other functions too [5]. ARs are crucial for the development and maintenance of the male sexual phenotype. Steroid molecule like TES either directly binds to the androgen receptor, or is converted by 5-alpha-reductase to DHT, which is a better agonist for androgen receptor activation [6]. TES is found to be the primary androgen receptor-activating hormone in the Wolffian duct, on the other hand dihydrotestosterone is the main androgenic hormone in the urogenital sinus, urogenital tubercle and hair follicle. Therefore it is believed that TES is mainly responsible for the development of primary male sexual characteristics, whereas DHT is responsible for secondary male characteristics [7].

AR is responsible for normal development and homeostasis of both male and female reproductive organs and their physiology. Mutation of certain residues of AR can have significant impact on AR activity. In fact, more than a thousand cases with pathogenic mutations affecting the human AR gene have been reported. These variations can be the cause of a dysfunctional receptor and lead to androgen insensitivity syndrome (AIS), which depending on the clinical phenotype is classified as complete (CAIS), partial (PAIS) or mild (MAIS). Moreover, it was reported that a large number of gain-of-function AR mutations have been associated with one of the leading causes of cancer death in men worldwide, castration-resistant prostate cancer. The rich clinical information on AR-related pathologies provides a detailed knowledge about the structure-function relationships for this transcription factor, and also for the other NRs [8].

Structurally, AR is similar to other NRs and consists of four segments, the N-terminal domain, followed by an almost strictly conserved DNA-binding domain (DBD), an interdomain linker or hinge, and a C-terminal ligand-binding domain (LBD). The LBD contains three binding surfaces which include the internal ligand-binding pocket (LBP) and two major solvent-exposed surfaces responsible for interaction with coregulators, activation function 2 (AF-2) and binding function 3 (BF-3) [11-14]. The structures of several full-length NRs were analyzed either by detailed X-ray crystallography, or through small-angle X-ray scattering and electron microscopy at lower resolution [8]. Previous modelling attempts of the full-length protein were guided by structural information on isolated domains and consideration of previously reported mutations. As the individual AR domains have autonomous functions (nuclear translocation, coactivator recruitment, DNA and ligand binding), several intra- and inter-domain interactions are critical for the integration of input and output signals required for proper AR functioning. Previous

experimental results suggest that changes made on one site of a nuclear receptor can have effect on another site of the receptor and this kind of effect which is known as an allosteric effect. Applying mutations on certain BF-3 residues and biophysical analysis of the structure might help us understand the allosteric control of various AR functions [9].

Here we experiment the crystal structure of the human AR LBD structure bound to its natural agonist, testosterone (TES) and discuss the unbinding pathways of TES from AR receptor and its mutants which might be critical in understanding the reasons behind the unbinding of TES from AR receptor and its mutants and how to prevent it. In order understand the unbinding pathways of TES multiple molecular dynamic (MD) and steered Molecular dynamic (SMD) simulations were performed on the AR LBD-TES complex and two AR mutant receptors. We examine if specific mutations on the BF-3 site would allosterically affect the unbinding of TES from the ligand binding pocket. The unbinding pathways were determined by using Caver 3.0. CAVER is a software tool which is commonly used to identify and characterize the transport pathways in static macromolecular structures. CAVER 3.0 implements new algorithms for the calculation and clustering of pathways. A trajectory from a molecular dynamics simulation serves as the typical input, while detailed characteristics and summary statistics of the time evolution of individual pathways are provided in the outputs. We performed SMD on the three systems and analyzed the data to make an assumption about the unbinding pathways of TES from the AR LBD-TES complex and the two AR mutant receptors.

CHAPTER II

LITERATURE REVIEW

Nuclear Receptors

Nuclear receptors are a family of ligand-regulated transcription factors that are activated by steroid hormones, such as testosterone, and various other lipid-soluble signals, including retinoic acid, oxysterols, and thyroid hormones. The implication of nuclear receptors (NRs) in a wide variety of complex biological processes and pathologies make them major pharmacological targets [10-12]. NRs are highly social proteins and multiple protein partnerships modulate their context- and time-dependent activities. NR modular architecture is behind this plethora of functions and reflects their common evolutionary origin. NRs are composed of an amino-terminal domain (NTD), a central DNA-binding domain (DBD), a hinge region, and a carboxyl-terminal ligand binding domain (LBD) [13]. The vast majority of structural and functional studies in this field have focused in structural characterization on isolated LBDs from different receptors. The LBD has a common conserved structure with the hormone nesting in its interior. The AF-2 pocket, a highly conserved protein-protein interaction site on the LBD surface, which recruits coactivators, has also been widely analyzed throughout this superfamily. The availability of a large number of LBD crystal structures proves the great functional versatility of this fold, however, it raises the question of how these proteins gain specificity or fine-tuning as numerous partner proteins are shared among NR subtypes [14].

Androgen Receptor Binding function-3 (AR BF-3)

Several lines of evidence show that the AF-2 groove of LBD may not be the only protein–protein interface on the surface of AR. Using a combination of X-ray crystallography and functional assays, another novel site was discovered on the human AR LBD surface and named as Binding Function (BF)-3 [9]. Previous studies with other NRs have also revealed protein-protein interacting surfaces distinct from AF-2, some of which overlap with AR BF-3. Structure-based sequence alignment of multiple NR LBDs proves that the BF-3 pocket is conserved among steroid receptors (SRs) and also present in other major NRs. Previous studies also suggest that a number of missense mutations that are within the vicinity of AR BF-3 pocket can lead to prostate cancer, infertility, and/or androgen insensitivity syndrome. Several mutations in the BF-3 site of other related NRs have also been linked with pathology or abnormal NR function in vitro.

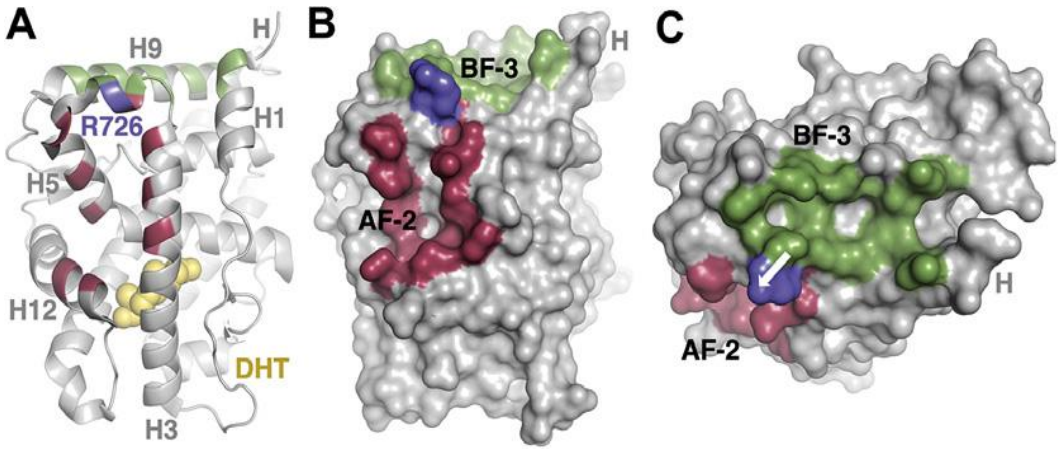

Figure 1. Androgen receptor LBD representation. (A) Illustration of AR LBD showing the location of the buried steroid in the ligand binding pocket (DHT, space filling model in yellow), and helices 1 (H1), 3 (H3), 5 (H5), 9 (H9) and 12 (H12) (grey). Residues lining the AF-2 pocket are highlighted in raspberry whereas, colored in green are BF-3 residues. The residue R726 present at the boundary between AF-2/BF-3 is shown in blue. (B) Space-filling model showing AR LBD surface (grey) and relative location of AF-2 and BF-3 pockets. (C) 90° rotation of AR LBD depicted in (B) to display BF-3 fully. H is the hinge region [9,14].

The BF-3 pocket in the AR was accidentally found by X-ray crystallography, in vitro transcriptional assays, and site-directed. mutagenesis [9,14]. BF-3 is reported to be concave and topographically close to but distinct from the AF-2 coactivator-binding groove on the LBD (Fig. 1). The BF-3 pocket, having similar size and depth to the AF-2 pocket, is solvent exposed and hydrophobic in nature, acts as a protein–protein interaction site (Fig.1). AR BF-3 is enclosed by several LBD-forming helices and resembles a rectangular shape with rounded corners [15].

Ligand Unbinding from Estrogen Receptor

Like Androgen receptors, Estrogen receptors ERα and ERβ are also members of the nuclear receptor (NR) superfamily and work as ligand-induced transcription factors assisting many of the estrogen actions in human body [16-17]. So, computational works done on that receptor was useful towards our study. Theoretical studies of ligand unbinding in NRs have been done with several MD methods such as steered molecular dynamics (SMD) [18] simulations locally enhanced sampling method (LES) [19], and random acceleration molecular dynamics (RAMD) [20] simulations. RAMD was applied to study the ligand unbinding event from ERs which enhances the MD sampling by addition of an external force to cross energy barriers. In this method bound ligand looks for its way out of the closed ligand-binding pocket, without any pre-knowledge of the pathways. Further pathway characterizations were obtained with the SMD method. Another software named CAVER [21] was also used to determine and validate the unbinding pathways of ligand from ER, which uses a probe sphere with a previously defined radius and a penalty function, based on the van der Waals size of protein atoms, to define the pathways .

The simulations were done in the presence of cofactors and with both agonist and antagonist ligands (17b-estradiol, genistein, and 4-hydroxytamoxifen). Their results showed that agonists or selective ER modulators can exit from the receptor through multiple pathways with minor effect on the receptor structure, whereas an antagonist requires larger conformational changes. Furthermore, they also concluded that a specific receptor-ligand combination might prefer an unusual pathway depending on the character and conformation of both parts [22].

Suggested Pathways for Dissociation of TES from AR Obtained by RAMD

A group of researchers at University of California (UC), Berkley used RAMD simulations [20] implemented in the NAMD 2.13 program suite [23] to investigate the potential dissociation pathways of nine ligands (ranging from agonists to strong antagonists) from the ligand binding site of AR. Since the ligands can potentially exit from a variety of channels, 80 different simulations were carried out to get good statistics of the ligand exit tunnels. In each of them, a random force was applied to accelerate the ligand movement. The initial force direction was randomly chosen by the algorithm. During the simulation, the force direction was retained if the ligand center of mass moved by certain distance in that direction. Agonist and antagonist ligands (17β-trenbolone, methyltrienolone, DHT, TES, p′, p-DDE (1,1-dichloro-2,2-bis(p-chlorophenyl) ethylene), neburon, vinclozolin, hydroxyflutamide, bicalutamide) were docked into the ligand binding pocket of AR. The main difference between the agonist and antagonist ligand binding is the presence or absence of H-bonds between the ligand and the receptor. The RAMD trajectories revealed the unbinding of testosterone through three different channels A, B, and C (Figure 2). A total of 62.50% of the egress trajectories were via channel A, another 26.25% were via channel B, and the remaining 11.25% were via channel C [24]. The results of this study will be used to compare the results of our study which is examining the exit pathways for TES form wild type AR and two mutant receptors with mutations in the BF-3 site.

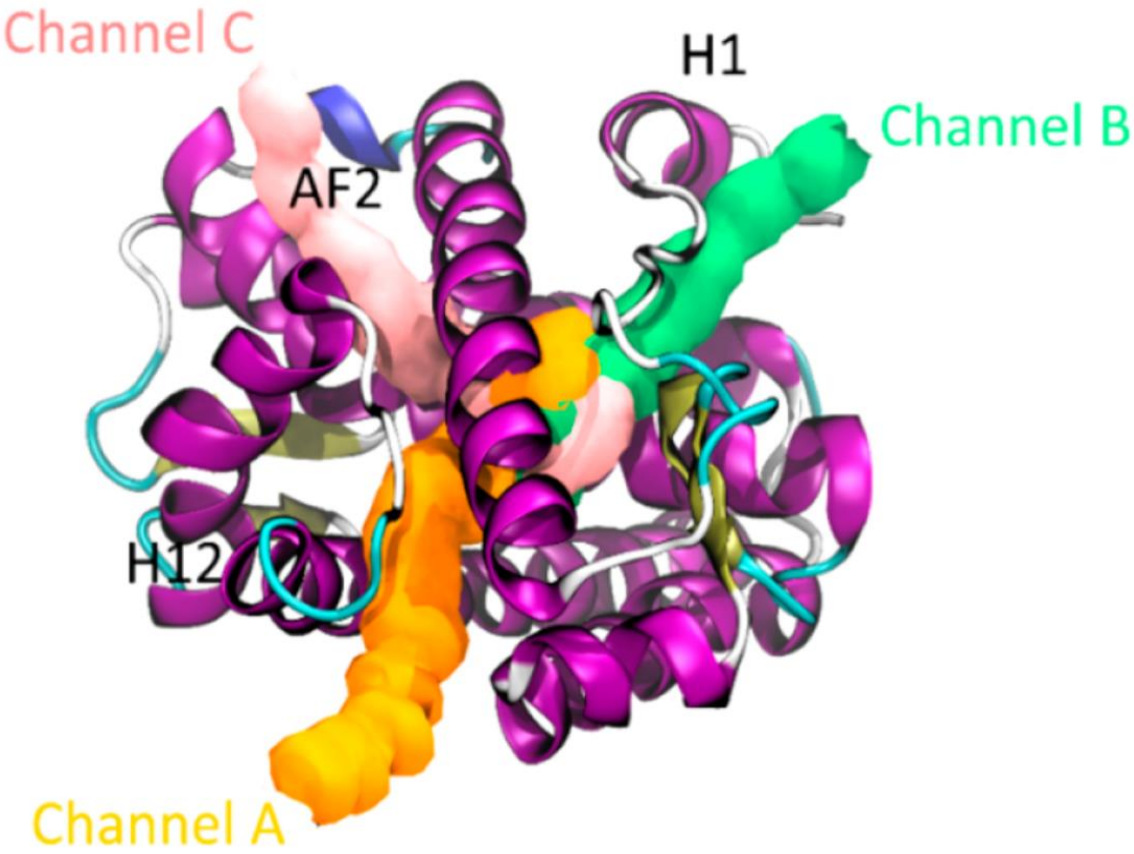

Figure 2. Suggested Pathways for dissociation of TES from AR. Pathways are shown in colors as Channel A, B, C obtained by RAMD [24].

Project Motivation

Recent experimental studies revealed that certain mutations in the BF-3 surface binding site can have an effect on the concentration of steroid required to activate the AR receptor (Table 1). The F673K mutation causes the AR to require a 670-fold increase of concentration of the steroid DHT in order to activate compared to the wild type (WT). On the other hand, the G724M mutation had no effect on AR activation, since the receptor required approximately the same concentration of steroid as the WT to activate, as shown in Table 1. For the rest of the mutants,

the DHT concentration change ranged from 6-fold to 19-fold. The pronounced effect of the F673K mutation and the lack of effect of the G724M mutation led us to select these two mutations for this study.

We hypothesized that if a mutant AR required a higher concentration of steroid to activate, such as the F673K mutant, that would be because the receptor is destabilized, and the steroid cannot bind as stably to the steroid binding site, and so a higher concentration of it is required to activate AR. This would imply that mutations in the BF-3 site that change the EC_{50} of DHT for AR have an allosteric effect on the steroid binding site. More specifically, we expected to see differences between the wild type receptor and the F673K mutant, in the interactions of TES with the amino acids in the steroid binding pocket. We also expected to see differences in the unbinding pathways of TES between the WT and the F673K mutant, such as which pathways are favored based on total force applied, maximum force, and unbinding time. Finally, we expected that the G724M mutant would not behave differently from the WT, since the mutant receptor required the same DHT concentration to activate [25].

Table 1. DNA vs EC_{50} concentration of DHT required to activate the receptor.

DNA	**EC_{50} of DHT**	**Fold Change of DHT Concentration**
WT	0.30 nM	-
F673K	201 nM	670
F673W	1.9 nM	6
G724R	2.80 nM	9
G724M	0.24 nM	1
L839D	5.6 nM	19

CHAPTER III

MATTERIALS AND METHODS

Preparation of the Systems

The initial AR LBD-TES complex that served as the control system, was constructed using the coordinates from the Protein Data Bank file ID 2AM9, which contains the AR LBD with TES (Fig. 4) bound in the steroid binding pocket. Hydrogens were added to AR LBD-TES complex using the Protein Preparation Wizard from Maestro (Schrodinger Software Suite) and an energy minimization calculation was performed using the Schrodinger module Macromodel to optimize their geometry. Two additional single-mutation AR LBD-TES systems were prepared by using Maestro to apply the mutations F673K (Fig. 3C) and G724M (Fig. 3D) in the BF-3 site of AR LBD-TES complex.

Each of the three receptor-ligand systems were solvated in a rectangular water-box with dimensions 75 Å x 67 Å x 79 Å using TIP3P water molecules [26] (11,936 water molecules) and ionized using NaCl to an ionic strength of 0.15M to simulate the cell cytosol. Figure 5 illustrates the AR LBD using ribbon representation the TES in space filling representation and the waters of the solvation box using point representation. Chemistry at Harvard Macromolecular Mechanics (CHARMM) force field parameters for TES were developed using the CHARMM [27] Generalized Force Field ref implemented through the ParamChem website [28-29] The CHARMM 36 force field ref was used for the receptor. All molecular dynamics simulations

were performed with the program NAMD [23] and the visualization software VMD [30] was used to visualize and analyze the resulting trajectories.

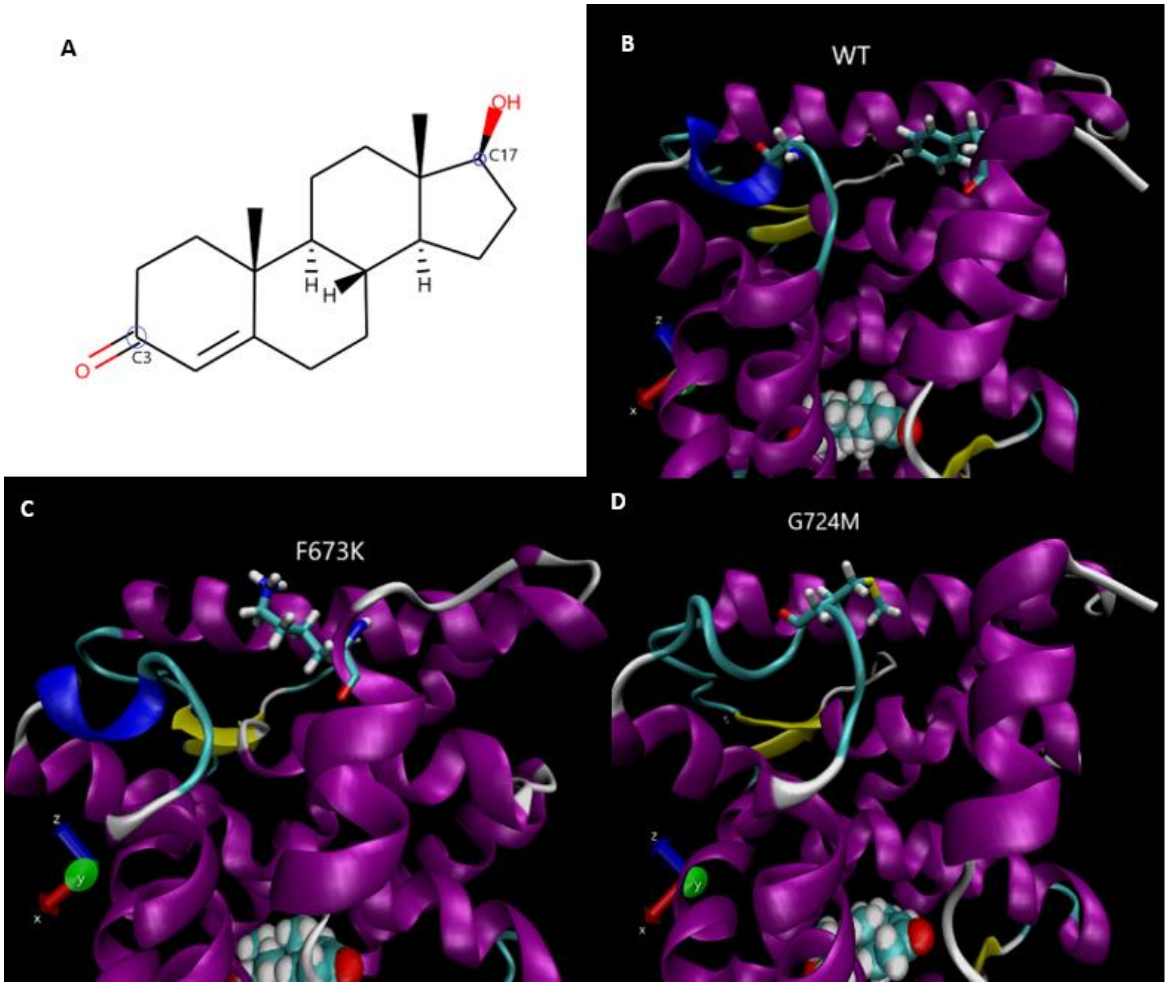

Figure 3. A represents a 2D structure of Testosterone molecule (TES). B illustrates the BF-3 site of WT. C and D illustrate the BF-3 site of the F673K and G724M mutants.

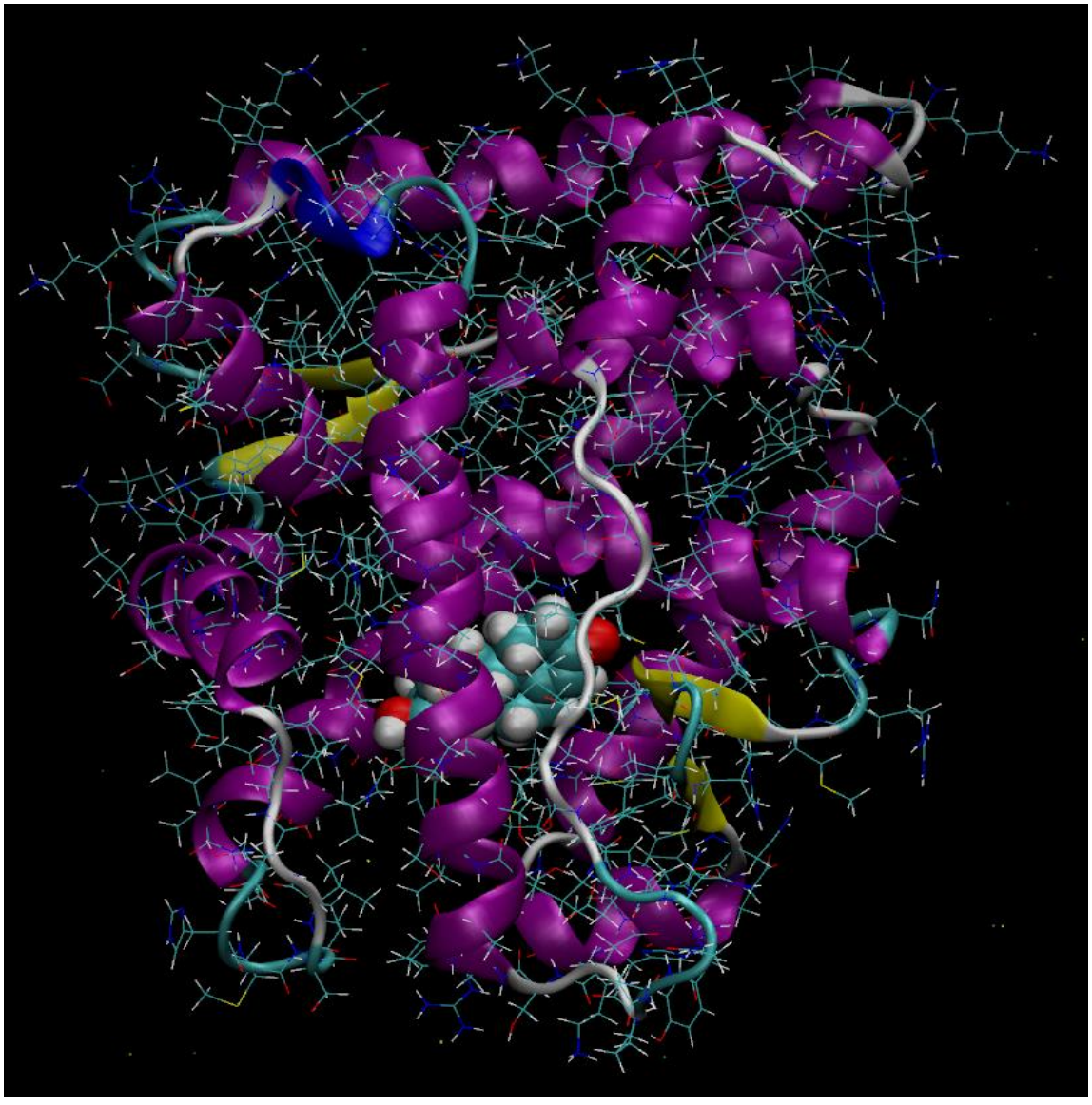

Figure 4. AR LBD-TES complex in ribbon representation displaying the side chains in line representation.

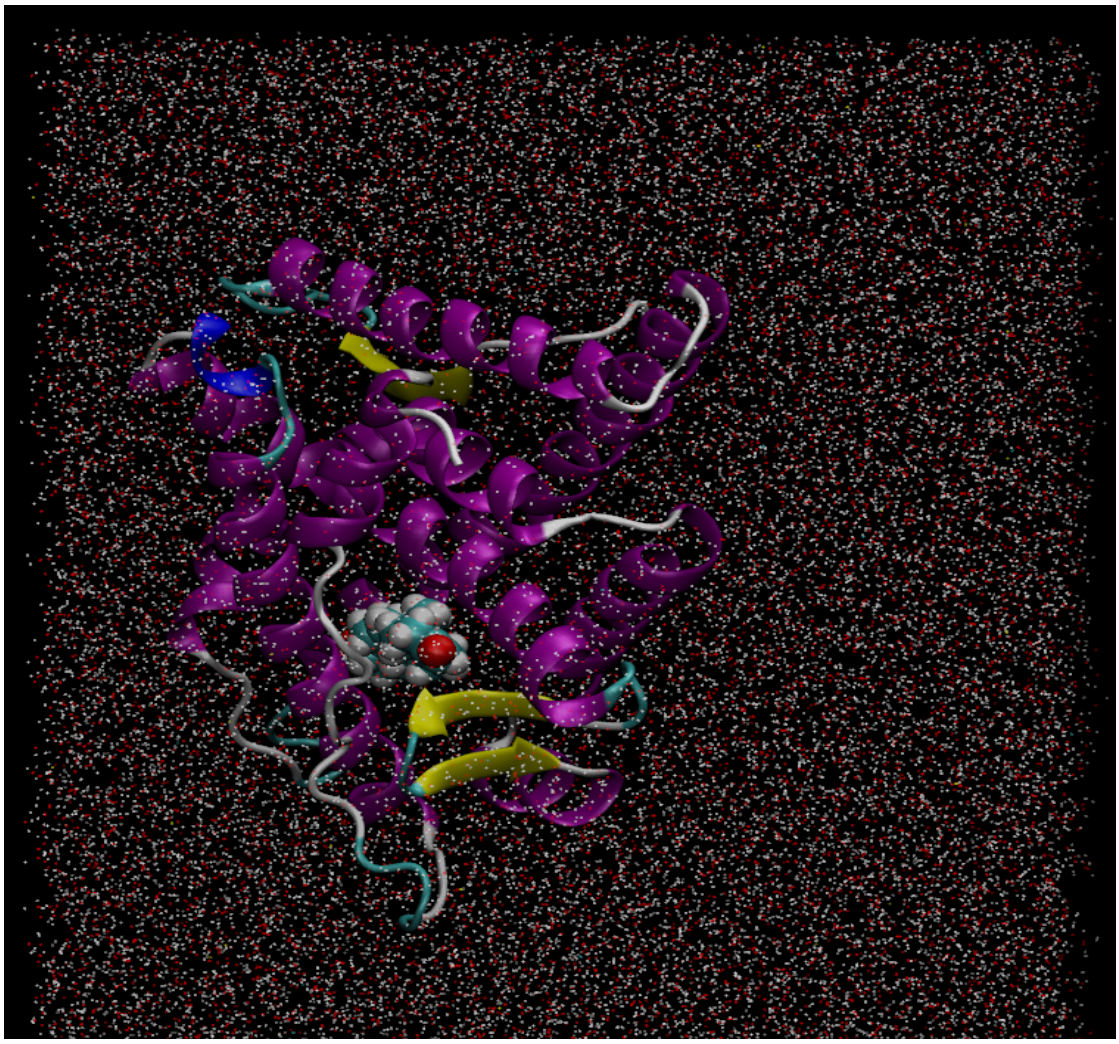

Figure 5. AR LBD-TES complex solvated in a water box showing the solvent in point representation.

Stochastic Dynamics Simulations Setup

We performed stochastic dynamics (SD) simulations on three previously systems and the structures obtained from these simulations were used as input for the Caver Analysis. For the SD simulations, the systems were initially minimized using the Polak-Ribier Conjugate Gradient method employing the OPLS2005 force field and the GB/SA solvation model. The systems were then simulated for 1 ns using SD simulations obtaining 100 random structures as output to be used for analysis. Both the minimization and SD simulations were performed using Macromodel.

Molecular Dynamics Simulations Setup

MD simulations were performed using the software NAMD. The non-bonded cutoff value for the force field parameters was set at 11.5Å with a switchdist set at 10Å. The three systems WT, F673K and G724M mutants, were minimized in three steps where we gradually reduced the constraints put on the protein backbone atoms and the TES atoms from 10 to 1 to 0 kcal/mol/$Å^2$ for a total of 12ps. After the minimization, the three systems were warmed up to 310K in 10K increments using Langevin Dynamics [31]. The Particle Mesh Ewald Sum [32] method was used to calculate the electrostatic interactions. The systems were then gradually equilibrated using the NPT ensemble at 310K and 1.01325 bar (1atm) pressure for 50 ns using a 2-fs timestep and rigid hydrogen bonds. After the systems equilibrated, another 10 ns of production run were carried out. For the equilibration and production runs the energy and structure information was collected every 10 ps. We analyzed the resulting simulations using VMD. We obtained the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) plots for the equilibration and production runs for all the systems. We also carried out

an analysis of the hydrogen bonding interactions between TES and four amino acids of the steroid binding pocket: N705, Q711, R752 and T877.

CAVER Setup

The software CAVER 3.0 was used to analyze the 100 frames that were obtained from the SD simulation for each of the three systems. The AR LBD-TES WT and mutant structures were stripped from water, ligand, and ions. After that, we requested for tunnels starting from the ligand binding pocket to a point outside the structure with a restricted grid resolution of 0.9 Å. The resulting pathways were analyzed for similarities, labeled and ordered according to the priority score provided by CAVER. The CAVER priority score is based on the length and width of the pathways. CAVER was also used to determine the residues lining the pathways.

Steered Molecular Dynamics Simulation Setup

Steered molecular dynamics (SMD) simulations are similar to MD simulations but an external constant or variable force is applied to an atom in a defined direction to pull that atom away from a refence atom or set of atoms. The two common types of SMDs are: constant velocity SMD and constant force SMD. For this project we employed constant velocity SMD to pull the TES molecule out the three AR LBD systems in the directions obtained from caver analysis. For this purpose, a dummy atom attached to the TES carbon atom 3 or 17 (Fig. 3A) was translated to a final position outside the receptor depending on the direction they were pulled. The velocity of the translation was set to be 0.035 Å/ps. The dummy atom was attached to the ligand atom that defines the starting position of the translation with a harmonic spring with force constant (k) of 4 kcal/mol/Å^2. The SMD simulations were performed at a constant temperature of 310 K. The force profile was calculated based on the following equation, $F(t) = k(x(t) – x_o(t))$

where k is the force constant of the harmonic spring, x(t) the position of the ligand atom attached to the spring and x_o (t) the position of the dummy atom at time t. The total force, F_{sum}, was calculated by adding up F(t) over the SMD trajectory and was reported in arbitrary units.

CHAPTER IV

RESULTS AND DISCUSSION

MD simulations of the WT and mutant receptor systems

All the AR LBD-TES systems, the WT AR and the two mutants, F673K and G724M, were built and equilibrated for 60 ns. The last 10 ns were used as the production run. A comparison of the RMSD for the three systems is reported in Figure 6. The RMSD calculation shows that all the systems stabilized after 50 ns (5000 steps) and were ready for the steered

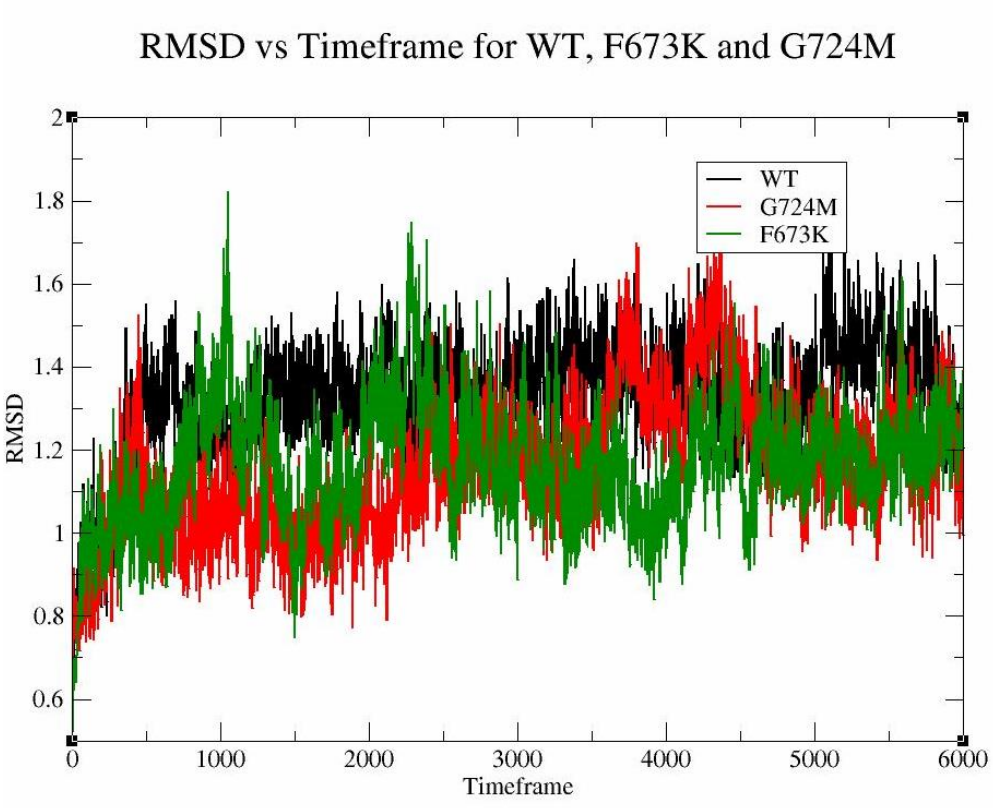

Figure 6. RMSD vs. time frame for the AR LBD-TES complexes

molecular dynamics simulations. The RMSD of the WT appeared to have flattened sooner compared to the RMSD of the two mutant receptor complexes.

For all the systems, TES was observed to be stably bound in the steroid binding site and maintain the hydrogen bonding interactions between the carboxyl oxygen in the 3 position and the amino acids T877 and N705, whereas Q711 and R752 did not maintain hydrogen bonding interactions the hydroxyl group on position 17 as shown in Table 2. These hydrogen bonding interactions have been observed in the crystal structure of AR used for this study. The data reported in Table 2 represent the percent occurrence of these four hydrogen bonds for the last 10 ns or 1000 frames of the equilibration time when all three receptor-ligand systems had stabilized. Table 2 shows that the frequency of hydrogen bonding between the ligand and the amino acid T877 was lower by 5.3% for the F673K mutant AR compared to both the WT and the G724M mutant. For the residue Q711 the hydrogen bonding occurrence with TES for the F673K system was also lower compared to both WT and the G724M. However, in the case of the residue N705, the hydrogen bonding interaction occurred more frequently in both the mutants compared to the WT AR. Finally, R752 did not make a direct hydrogen bond with TES during the last 10 ns of the simulations. During the equilibration, the hydrogen bond between R752 and the hydroxyl in position 17 of TES was broken and R752 was solvated in the water for the majority of the simulations.

Disruption of the hydrogen bonding interactions between TES and the four amino acids in the mutant AR receptors would suggest a destabilization of the steroid binding pocket. Since for each of the mutant receptors, one that required increased amounts of DHT to activate and one that did not, two of the three hydrogen bonds were observed less frequently but the third one increased in frequency compared to the WT, it was hard to draw the conclusion that the pocket

got destabilized. Additionally, this study did not account for any differences in the other types of interactions, especially the Van der Waals (hydrophobic) interactions, that occur between TES and the amino acids of the binding pocket. However, it was clear that there was a change in the pattern of the hydrogen bonding interactions, so the steroid binding pocket was affected in an allosteric manner by the two mutation in the BF-3 site.

Table 2. Hydrogen bonds between AR LBD and TES.

Systems	T877	N705	Q711	R752
WT	83.3%	64.2%	1.4%	0.0%
F673K	78.0%	83.2%	0.0%	0.0%
G724M	83.3%	84.1%	0.9%	0.0%

The RMSF of the AR LBD amino acids is reported in Figure 7. The termini since they are loose, tend to have higher RMSF compared to the rest of the protein. Also, the loop regions are more flexible leading to higher RMSF values as can be seen by the large peaks around amino acids. It is also expected that the RMSF of the terminal residues will be higher compared to the other residues. It was observed that the BF-3 site mutations did have an effect on the RMSF of the protein. Both mutants seem to differ from the WT in terms of RMSF however, the F673K mutant was observed to have been affected more compared to G724M mutant. The residues from 680 to 686, which belong to the loop between helix (H) 1 and H3 (H1/H3 loop), were fluctuating more for the WT compared to F673K and G724. The residues from 697 to 699 are from the bottom part of H3 and they were fluctuating more in case of the WT compared to G724M and F673K. H3/H5 loop amino acids 722 to 730, which are part of the BF-3 site, fluctuate more for the F673K mutant. This is because the lysine of the F673K is next to the residues 722 to 730 and is partially in contact with them since it is more exposed to the surface than the phenylalanine of

the WT. The residues from 756 to 760 that were also fluctuating about the same amount for all the three systems belong to the loop between H5 and H6. The residues from 770 to 780 belong to H6 which was moving most for the WT compared to the G724M and F673K mutants, with F673K fluctuating the least. The residues from 780 to 800, which are part of the bottom of H7 and the H7/H8 loop, were moving much less for the F673K mutant compared to G724M and the WT. Again, the residues from 816 to 825, which are part of the H8/H9 loop, fluctuate more for WT compared to the two mutants. Finally, the amino acids 860 to 905 belonging to H10 and the loop between H10 and H12, fluctuated more for the WT compared to the two mutants. The RMSF analysis indicates that the two mutant receptors fluctuated similarly to the WT with few regions where the mutant receptors fluctuated less than the WT. This implies that the receptor was not affected in a major way by the mutations.

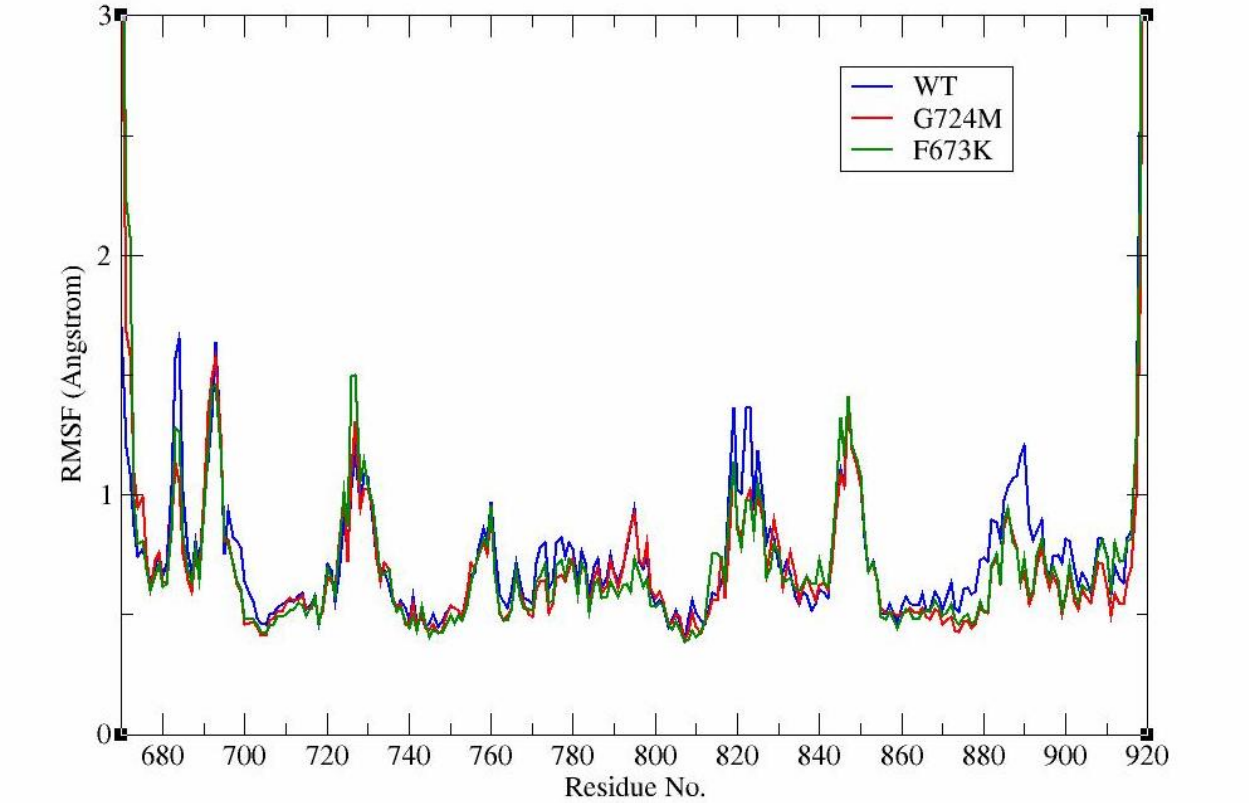

Figure 7. Comparison of RMSF for the WT AR LBD-TES and the two mutant systems.

Unbinding pathways found by CAVER

The CAVER 3.0 software was used to analyze all three receptor-ligand systems. The CAVER analysis identified a total of eight pathways which were very similar among the three systems. Some of pathways were very lengthy and very similar to one another so, we clustered them together and ended up with four distinct pathways which were approximately the same for the WT and the two mutant receptors. The exit pathways were named A, B, C, and D (Figures 8 and 9). The priority score, length and width of the pathways are reported in Table 3. The priority score is calculated by taking into account the length and width of the pathways as well as the

frequency of its occurrence in the total number of frames. The shorter the length of the pathway and the wider the radius, the more probable is the pathway.

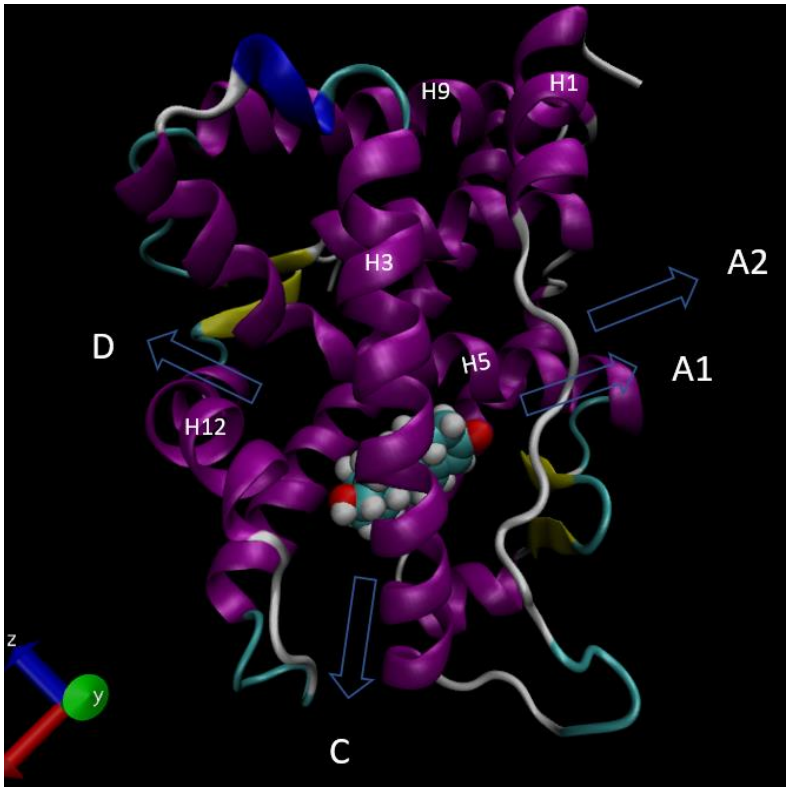

Figure 8. A schematic of the five pathways obtained by CAVER analysis.

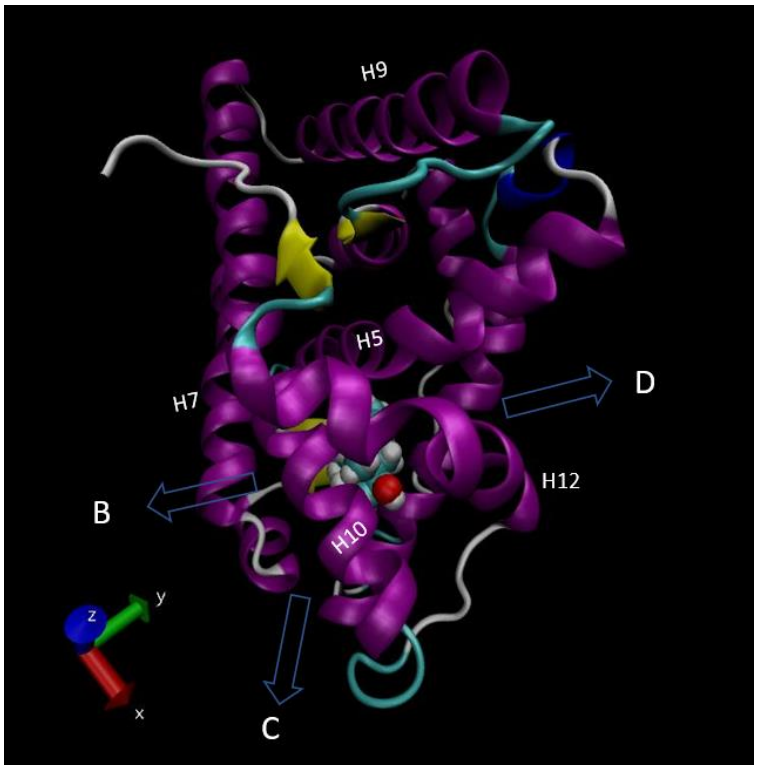

Figure 9. Rotation of the receptor in Figure 8 by 90° on the vertical axis.

The pathways A, C, and D matched with the pathways obtained by RAMD simulations done on the WT AR LBD by a group of researchers at UC Berkeley [24]. Tunnel A had two similar exits, one toward the left side of the H1/H3 loop connecting H1 and H3 (between H1/H3 loop and H3), whereas the other one toward the right side of the same H1/H3 loop (between the H1/H3 loop and H5). They were named A1 and A2 respectively (Figures 8 and 9). The tunnel that comes out through the gap between H7 and H10 was named pathway B. The tunnel exiting at the lower part of the receptor through the gap between H7, H11 and H3 was named pathway C. Finally, the tunnel that exited through the gap between H12, H5 and H3 was named pathway D. Based on the number of times they appeared and their priority score, pathways A and C appear to be favored in all cases. Pathway B was favored mainly by the WT and to a lesser extent

by the F673K mutant AR. Pathway D was the least favorite one for both WT and F673K. This should be due to the relatively large length and bigger width of the pathway.

Table 3. A quantitative summary of the tunnels obtained from caver analysis of the three systems ordered according to their priority scores.

System	Pathways	Priority	Number of snapshots (out of 100)	Average Length (Å)	Average radius (Å)	Maximum radius (Å)
	A	0.54973	95	17.198	1.190	1.62
WT	B	0.44865	75	13.497	1.147	1.75
	C	0.29096	52	14.466	1.054	1.44
	D	0.23793	56	18.175	1.036	1.48
	C	0.45611	77	15.295	1.246	2.03
F673K	A	0.37314	68	16.363	1.026	1.36
	B	0.25275	47	15.648	1.042	1.35
	D	0.193	43	24.973	1.023	1.18
	C	0.52524	90	12.743	1.079	1.44
G724M	A	0.50811	96	19.600	1.153	1.55
	D	0.2535	35	15.627	1.104	1.54
	B	0.01016	5	20.046	0.979	1.03

The residues present in the lining of the tunnels are reported in Table 4. From the RMSF graph for all the systems, we saw that the residues E681 and P682 present in the linings of tunnel A fluctuated considerably compared to other residues. The residue F764 is present in both tunnel B and C and fluctuated considerably based on the RMSF plot. On the other hand, the residue M895 present in the lining of tunnels B, C and D was also observed to be very flexible. Which explains why we got a high percentage of a caver tunnels in those directions.

Table 4. Residues that are lining the unbinding pathways of the WT AR LBD .

Tunnels or pathways	Residues lining the tunnels
A	681 GLU, 682 PRO, 704 LEU, 707 LEU, 708 GLY, 711GLN, 741TRP, 742MET, 745MET, 746 VAL, 748 ALA, 749 MET, 752 ARG, 764 PHE, 808 LYS, 873 LEU, 705 ASN, 683 GLY, 744 LEU.
B	701 LEU, 704 LEU 742 MET, 764 PHE 780 MET 783GLN 873 LEU 876 PHE 705 ASN, 708 GLN, 741 TRP, 877 THR, 895 MET
C	701 LEU, 704 LEU 742 MET, 764 PHE, 780 MET, 787 MET, 873 LEU, 876 PHE, 877THR, 880 LEU, 895 MET 741TRP, 697 PHE, 746 VAL
D	704 LEU, 705 ASN, 708 GLY, 712 LEU, 741 TRP, 742 MET, 745 MET, 764 PHE, 873 LEU, 895 MET, 898 ILE, 707 LEU, 711 GLN, 787 MET, 746 VAL

SMD simulations results

The force vs. timestep graphs for the WT AR LBD-TES and the two mutant AR systems for the unbinding paths are reported in Figures 10 through 14. The data obtained from force profile analysis are reported in Table 5. The force required to pull TES out of the steroid binding pocked increases at the beginning and can show one or more peaks where TES is blocked by amino acids on its way out of the pocket. Once TES has exited the AR LBD the force is reduced and does not change much. From the force profiles of the SMD simulations, a maximal force value (F_{max}) was extracted for each pathway for the three different AR LBD-TES systems. The position where F_{max} is observed indicates the bottleneck of the unbinding trajectory. SMD simulations are nonequilibrium simulations due to the presence of a small magnitude of external force exerted on the system. The simulated SMD pathways showed large force variations within each pathway (Table 5).

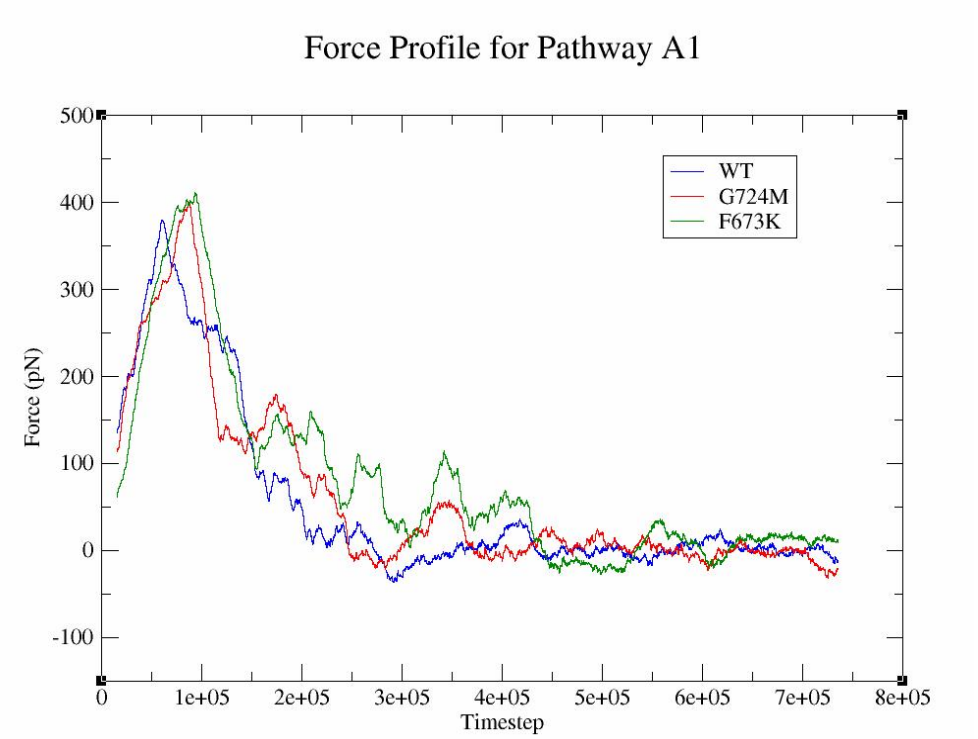

Figure 10. Force vs Timestep graph for the three systems for tunnel A1.

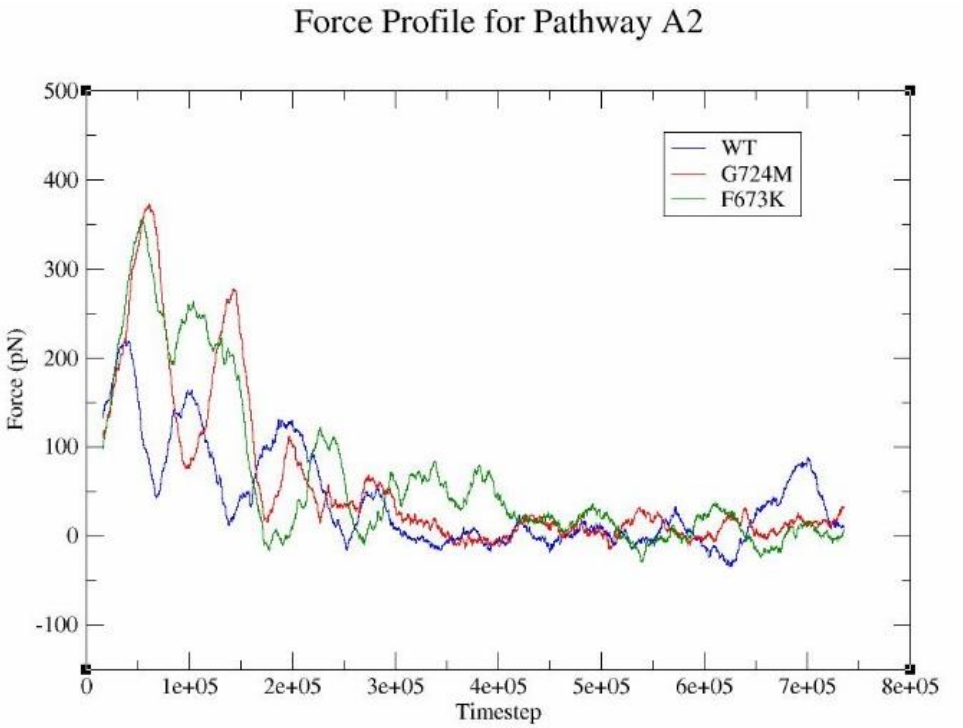

Figure 11. Force vs Timestep graph for the three systems in A2 direction.

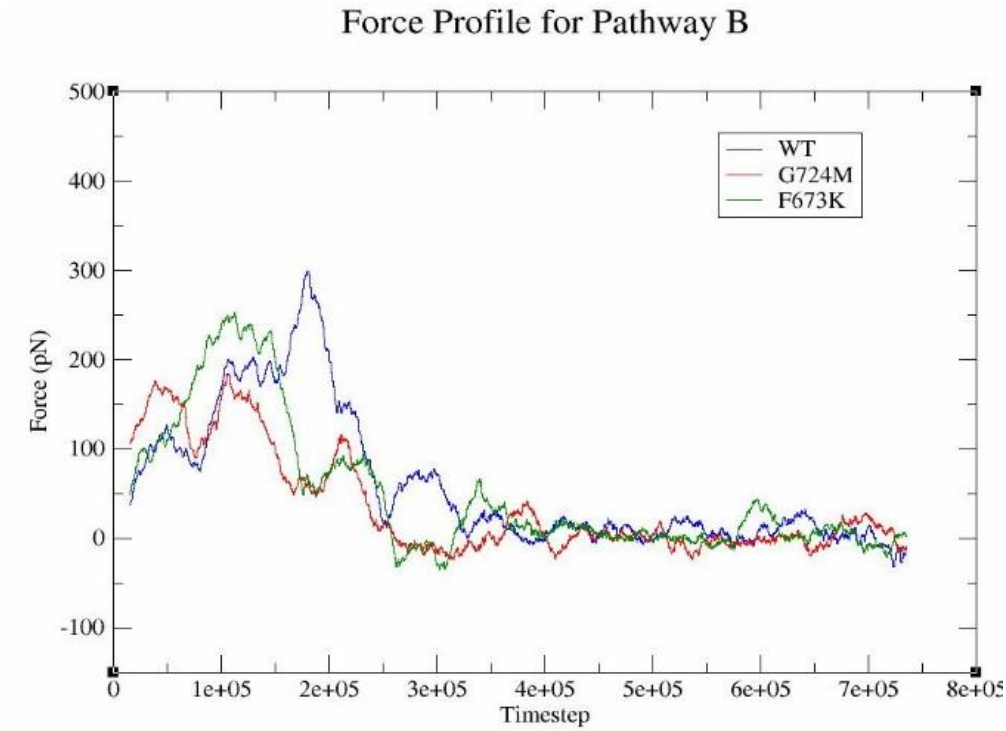

Figure 12. Force vs Timestep graph for the three systems in the B trajectory

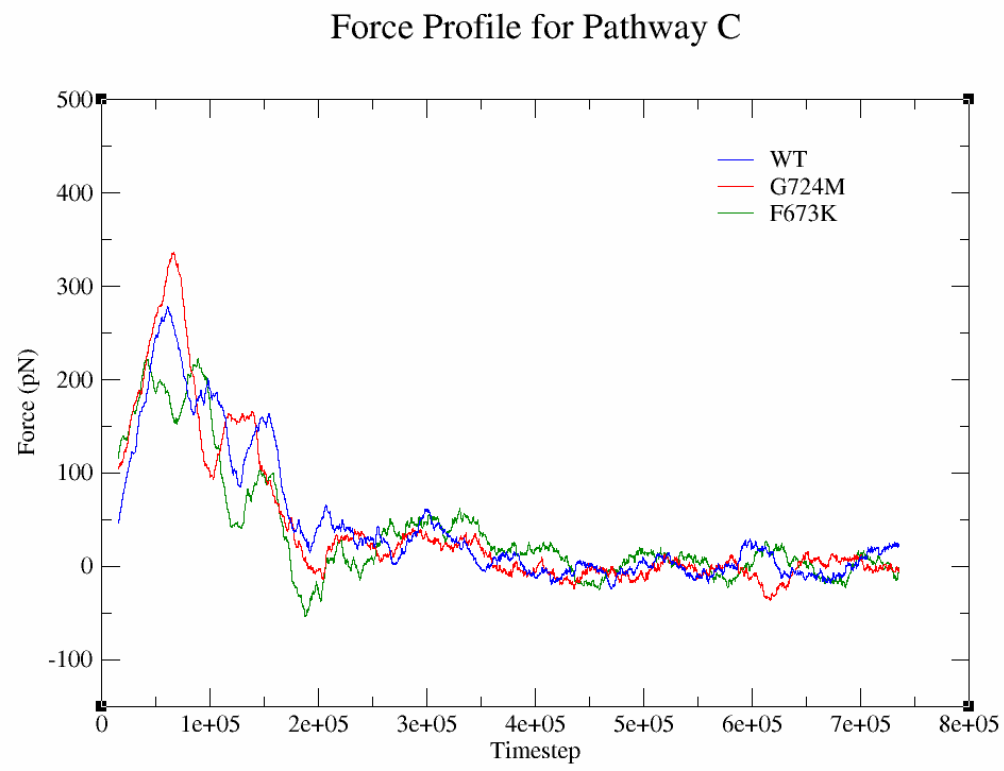

Figure 13. Force vs Timestep graph for the three systems in C trajectory.

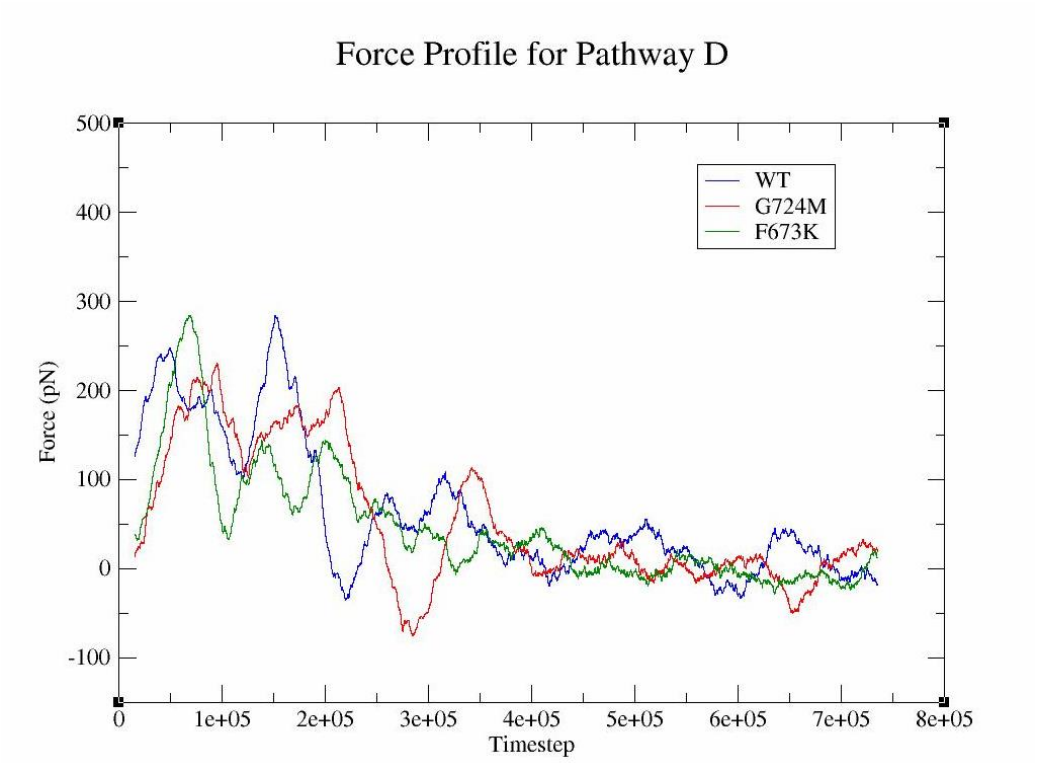

Figure 14. Force vs Timestep graph for the three systems in D trajectory.

To get an approximate measure of the total energy required for unbinding, we recorded the force output every 10 ps and we obtained the sum of the SMD force (F_{sum}) along the trajectory. A quantitative comparison of the force profiles indicates that some pathways are less favorable than others and the preference for pathways also varies depending on the receptor-ligand systems. It was observed that, the force F_{max} was less than 747 pN for all cases (all pathways and systems). For the WT, Pathway A1 and D showed both a high F_{max} (table 5) and a large F_{sum}, whereas the Pathways A2 and C showed a smaller F_{sum} and a smaller F_{max}. This observation implies that exit pathways A2 and C are the most favorable for the WT AR. This conclusion also is in agreement with the results previously reported after performing RAMD simulations on AR receptor [24]

The force profile analysis for the two mutants were significantly different from the WT. For the F673K mutant, unbinding pathways A1 and A2 had both high F_{max} and large F_{sum}, whereas the pathways B, C and D had a smaller F_{sum} compared to A1 and A2 and thus are better candidates. However, since pathways B and D require more time to exit the system and have higher F_{sum} than C, they are less favorable compared to pathway C. Similarly, in case of the G724M mutant, the pathways B, C, D were more favorable based on the value of the F_{sum} for all the pathways. Since pathway C required the least amount of time to exit and had the lowest F_{sum}, it is more favorable compared to the other pathways. From this information, we can conclude that overall, the unbinding pathway C was favored by all the three receptor-ligand systems.

Table 5. Results of SMD simulation for all the systems

Systems	Pathways	Time (ps)	F_{max} (pN)	F_{max} time (ps)	F_{avg} (pN)	F_{sum}
	A1	584	713	139	140	4095230
	A2	669	518	75	80	2664576
WT	B	671	661	370	118	3944603
	C	570	655	219	102	2920465
	D	761	663	291	112	4273548
	A1	690	704	160	155	5373593
	A2	723	677	132	120	4327099
F673K	B	669	608	90	98	3271854
	C	494	731	167	89	2199008
	D	746	658	122	87	3237182
	A1	687	711	180	130	4454130
	A2	681	746	111	116	3955855
G724M	B	604	626	210	108	3257839
	C	535	751	158	112	2990491
	D	815	626	210	92	3265546

Visual Observations on VMD

Visual inspection of the results of the SMD trajectory for the WT suggests that the pathway A1(Fig. 15A) was blocked mainly by the residues Q711, L707, V685, and P682. Q711 (a polar residue) forms a strong hydrogen bond with carbonyl oxygen of the TES molecule holding it in place. Once the bond is broken the force applied to TES to lead it out of the steroid binding pocket starts decreasing. A similar situation is observed for the A2 (Fig. 15B) exit pathway where Q711 was seen to partially block the way of the ligand causing a high F_{max} (Table 5). Another residue which initially assists the exit of the ligand by forming a hydrogen bond with TES, but later blocks the way of the ligand as it progresses through the A2 exit pathway is R752.

For the exit pathway via tunnel B (Fig. 16), it was observed that the polar residues E872, Q783, and R786 directly blocked the way of the ligand and formed hydrogen bonding interactions with the hydroxyl oxygen of TES causing increase in the force needed to pull TES out of the binding pocket. The hydrophobic residues F876 and M780 were also partially blocking the way which lead to a high F_{sum}. For the exit pathway C (Fig. 17) only hydrophobic residues like F876, F697 and L880 were blocking the way. The polar residue S782 was initially assisting the ligand exit by attracting the hydroxyl oxygen of TES but then partially blocked the way of the ligand with its steric bulk. Finally, for pathway D (Fig. 18), the residues M895, I898, I899 of H12 and L712 were in the way of the ligand. The ligand was observed to perform a rotation (clockwise) while trying to exit in this way.

After observing the SMD trajectories of the three systems we saw that ligand exit through pathway C happened in a very similar way for all the systems. Whereas tunnel A2 was observed to be blocked by more residues in case of the two mutants compared to the wild type. For the

F673K mutant AR (Fig. 19A) the A2 path was seen to be blocked by R752, Q711 and V684 and for the G724M mutant AR (Fig. 19B) it was blocked by M745, A748, R752, and Q711.The total force required to exit towards pathway C was similar for WT and the mutant G724M whereas it was much lower for F673K which supports the hypothesis that the receptor gets destabilized allosterically by the mutation. However, for both pathway B and D, we saw a reduction in the amount of force required to pull the ligand out of the systems for the two mutants. For pathway A1, the F_{sum} required to unbind TES from F673K was much higher compared to that required for the WT while for the G724M mutant is was only marginally higher. Higher F_{sum} means that it will be harder for the ligand to exit in that direction and hence less likely to happen. Overall, different unbinding pathways are favored for the mutant receptors compared to the WT supporting the hypothesis that the mutations will have allosteric effects on the unbinding pathways for AR.

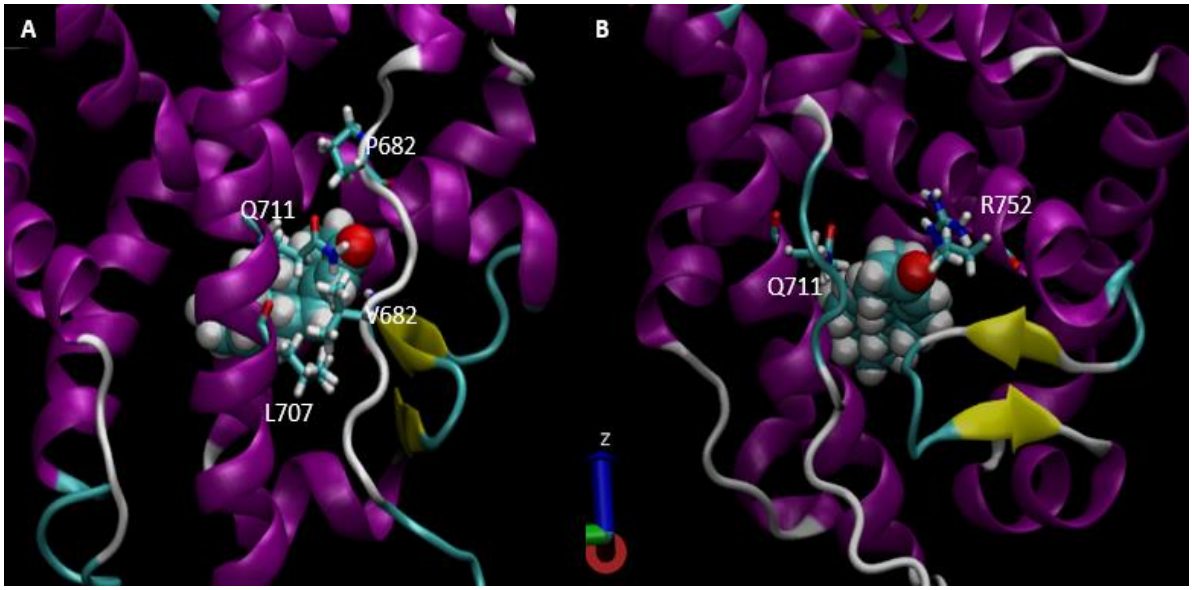

Figure 15. A illustrates the residues blocking pathway A1 for the WT. B illustrates the residues blocking the pathway A2 for the WT

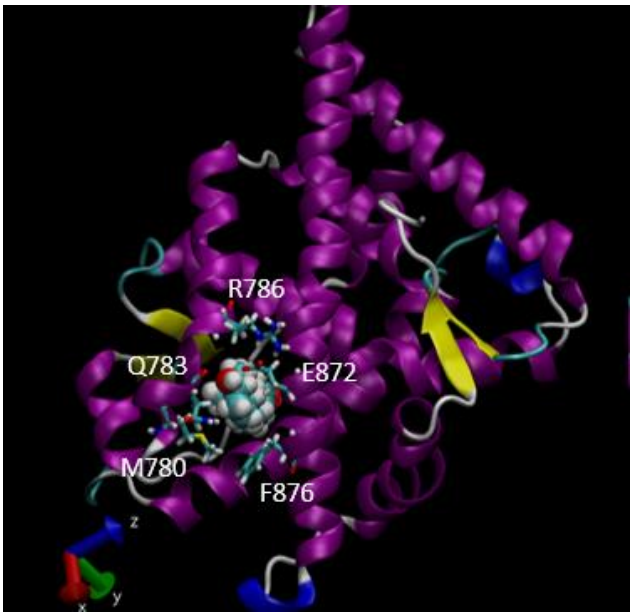

Figure 16. Residues blocking pathway B for the WT

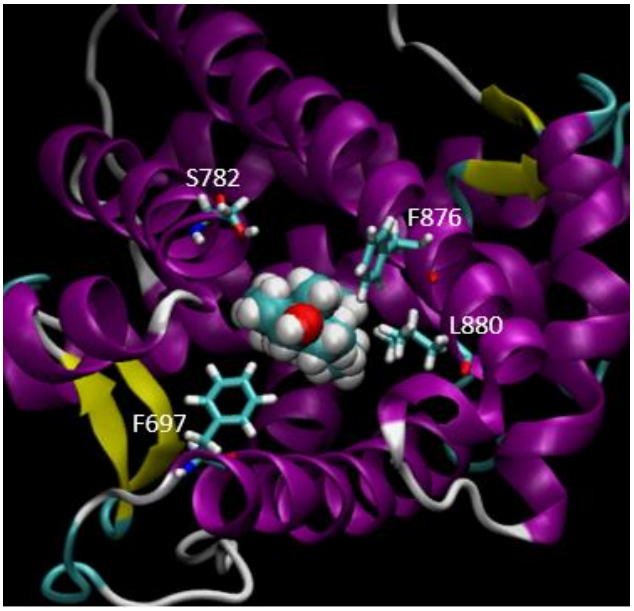

Figure 17. Residues blocking pathway C for the WT

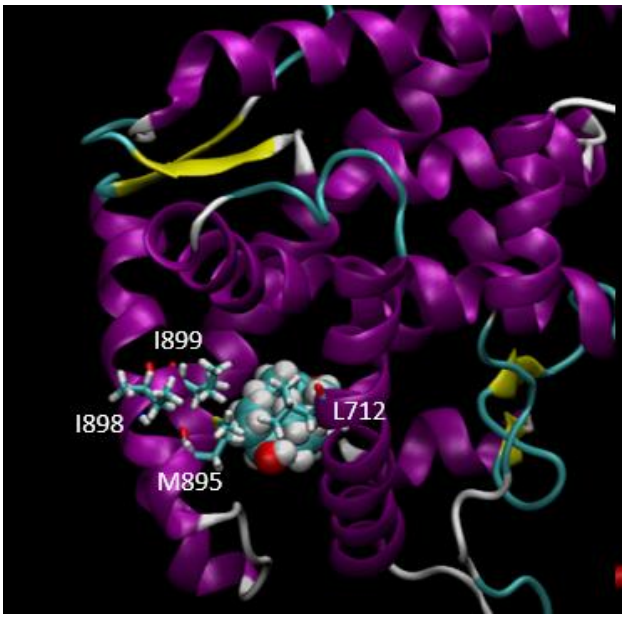

Figure 18. Residues blocking pathway D for the WT

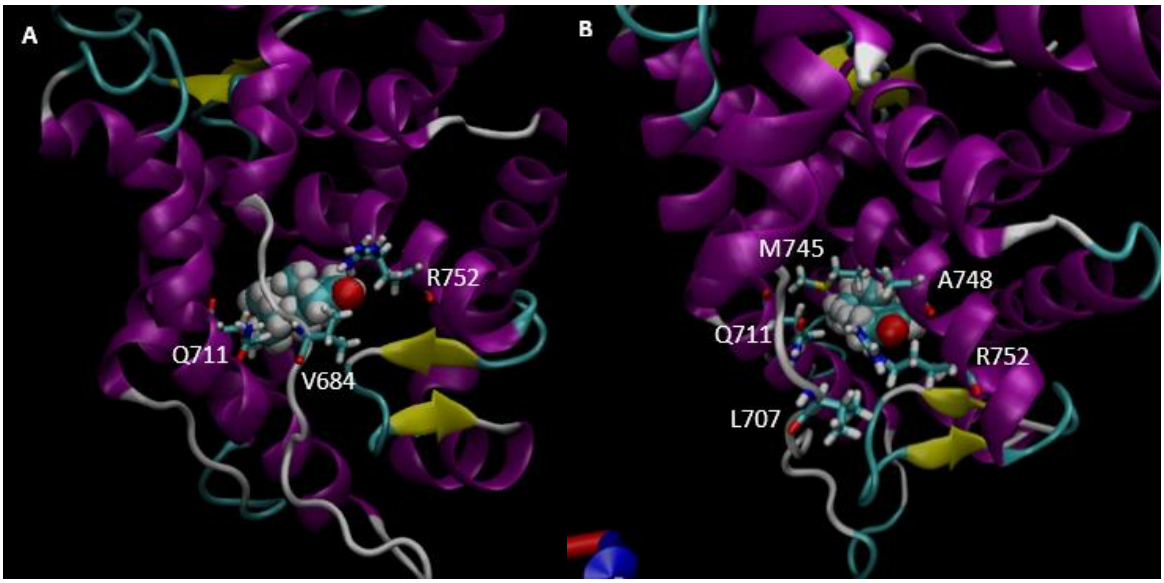

Figure 19. A illustrates the residues blocking pathway A2 for F673K. B illustrates the residues blocking pathway A2 for G724M mutants.

CHAPTER V

CONCLUSION

The theoretical studies presented here showed that both BF-3 binding site mutations examined, and not only F673K as expected, had an allosteric effect on the AR structure and the steroid binding pocket. The mutations did not have a large effect on the stability of the receptor. In fact, the mutant receptors showed lower RMSF in two helical regions, H6 and H10 as well as in the H8/H9 and H10/H12 loops. However, the mutations only induced small changes in the interactions of TES in the steroid binding pocket, specifically the hydrogen bonds between the residues T877, N705 and Q711 that interact with the carbonyl oxygen and the hydroxyl groups of TES. The percent occurrence of these hydrogen bonds changed compared to the WT but did not decrease drastically, as had been expected for the F673K mutant. Instead, the pattern of hydrogen bonding changed. The decreased occurrence of one of the hydrogen bonds was accompanied by an increased occurrence of one of the others.

The unbinding pathways of TES were affected by the mutations as well. Unbinding pathway A2 was preferred only by the WT, whereas pathway C was preferred by all three receptors. Pathways B and D required lower total force for the two mutants compared to the WT. Pathway D would be blocked when a coactivator is recruited at the AF-2 site and the A2 path is blocked when AR receptors dimerize to carry out their function. Therefore, we propose that pathway C is the most likely unbinding pathway for a ligand binding in the steroid binding site. Overall, we conclude that the binding trajectories were affected by the mutations on BF-3 site.

The results obtained from this study help us get a better understanding of how mutations in the BF-3 surface binding site can affect the AR structure and interactions between the steroid and residues in its binding pocket. Since the BF-3 binding site is a site of allosteric control for AR function, insights from this study can potentially guide the design of compounds that can bind to the BF-3 site and modulate AR function.

REFERENCES

1. Huang, P., Chandra, V. & Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu. Rev. Physiol. 72, 247–272 (2010).
2. Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR & the big bang. Cell 157, 255–266 (2014).
3. Roy AK, Lavrovsky Y, Song CS, Chen S, Jung MH, Velu NK, Bi BY, Chatterjee B (1999). Regulation of androgen action. Vitamins & Hormones. **55**. pp. 309–52. doi:10.1016/S0083-6729(08)60938-3. ISBN 978-0-12-709855-5. PMID 9949684.
4. Mooradian AD, Morley JE, Korenman SG (February 1987). "Biological actions of androgens". Endocrine Reviews. **8** (1): 1–28. doi:10.1210/edrv-8-1-1. PMID 3549275.
5. Heinlein CA, Chang C (October 2002). "The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions". Molecular Endocrinology. **16** (10): 2181–7. doi:10.1210/me.2002-0070. PMID 12351684.
6. Davison SL, Bell R (April 2006). "Androgen physiology". Seminars in Reproductive Medicine. **24** (2): 71–7. doi:10.1055/s-2006-939565. PMID 16633980.
7. Sinisi AA, Pasquali D, Notaro A, Bellastella A (2003). "Sexual differentiation". Journal of Endocrinological Investigation. **26** (3 Suppl): 23–28. PMID 12834017.

8. Nadal, M., Prekovic, S., Gallastegui, N. et al. Structure of the homodimeric androgen receptor ligand-binding domain. Nat Commun **8,** 14388 (2017).
9. Este´banez-Perpin˜a´, E. et al. A surface on the androgen receptor that allosterically regulates coactivator binding. Proc. Natl Acad. Sci. USA 104, 16074–16079 (2007).
10. Gao, W., 2010. Androgen receptor as a therapeutic target. Adv. Drug Deliv. Rev. (Epub ahead of print).
11. Gao, W., Bohl, C.E., Dalton, J.T., 2005. Chemistry and structural biology of androgen receptor. Chem. Rev. 105, 3352–3370.
12. Sadar, M., 2011. Small molecule inhibitors targeting the ‘‘achilles’ heel’’ of androgen receptor activity. Cancer Res. 71, 1208–1213.
13. Steinmetz, A.C., Renaud, J.P., Moras, D., 2001. Binding of ligands and activation of transcription by nuclear receptors. Annu. Rev. Biophys. Biomol. Struct. 30, 329–359.
14. Estebanez-Perpina, E., Moore, J.M., Mar, E., Delgado-Rodrigues, E., Nguyen, P., Baxter, J.D., Buehrer, B.M., Webb, P., Fletterick, R.J., Guy, R.K., 2005. The molecular mechanisms of coactivator utilization in ligand-dependent transactivation by the androgen receptor. J. Biol. Chem. 280, 8060–8068.
15. Buzón V, Carbó LR, Estruch SB, Fletterick RJ, Estébanez-Perpiñá E. A conserved surface on the ligand binding domain of nuclear receptors for allosteric control. Mol Cell Endocrinol. 2012;348(2):394-402. doi:10.1016/j.mce.2011.08.012.
16. Green S, Walter P, Kumar V, Krust A, Bornert JM, Argos P, Chambon P. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature 1986;320:134–139.

17. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A 1996;93:5925–5930.
18. Kosztin D, Izrailev S, Schulten K. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. Biophys J 1999;76:188–197.
19. Blondel A, Renaud JP, Fischer S, Moras D, Karplus M. Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. J Mol Biol 1999;291:101–115.
20. Carlsson P, Burendahl S, Nilsson L. Unbinding of retinoic acid from the retinoic acid receptor by random expulsion molecular dynamics. Biophys J 2006;91:3151–3161.
21. Petrek M, Otyepka M, Banas P, Kosinova P, Koca J, Damborsky J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinformatics 2006;7:316.
22. Burendahl S, Danciulescu C, Nilsson L. Ligand unbinding from the estrogen receptor: a computational study of pathways and ligand specificity. Proteins. 2009;77(4):842-856. doi:10.1002/prot.22503.
23. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. J. Comput. Chem. 2005, 26, 1781−1802.
24. Ettayapuram Ramaprasad Azhagiya Singam, Phum Tachachartvanich, Michele A. La Merrill, Martyn T. Smith, and Kathleen A. Durkin. Structural Dynamics of Agonist and Antagonist Binding to the Androgen Receptor. The Journal of Physical Chemistry B 2019 123 (36), 7657-7666

25. Unpublished data.

26. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79, 926−935.

27. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

28. CGenFF: K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell Jr., J. Comput. Chem. 2010, 31, 671-690.

29. ParamChem interface: www.paramchem.org

30. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. J. Mol. Graphics 1996, 14, 33−38.

31. Ruslan L. Davidchacka), Richard Handel, and M. V. Tretyakov. Langevin thermostat for rigid body dynamics. The Journal of Chemical Physics 130:23

32. Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98:12, 10089-10092

BIOGRAPHICAL SKETCH

Mr Muniruzzaman Chowdhury attended the University of Texas Rio Grande Valley from August 2018 to August 2020. He did his research in computational chemistry under the supervision of Dr. Evangelia Kotsikorou. He received an MS Degree in Chemistry from University of Texas Rio Grande Valley in August 2020. He can be contacted via email using the following email addresses: muniruzzaman.chowdhury01@utrgv.edu Or mchy94@gmail.com.