Identification of Novel Androgen Receptor Inhibitors for the Treatment of Castration Resistant Prostate Cancer

by

Ravi Shashi Nayana Munuganti

MSc. (Hons.), Osmania University, India, 2005

A THESIS SUBMITTED IN PARTIAL FULLFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies
(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2015

Abstract

Interest in developing androgen receptor (AR) inhibitors with novel mechanism of action for the treatment of prostate cancer (PCa) is on the rise since the commercial anti-androgens (including recently approved drug, Enzalutamide) face clinical limitations. Current therapies fail over a period of time because they all target mutation-prone androgen binding pocket on AR to which the receptor has already developed effective resistance mechanisms. Hence, there is a pressing need for novel therapeutics that inhibit the AR through alternative modes of action.

To address this problem, we have used *in silico* drug design methodology to create new drugs that act on an entirely different site on the AR, a recently identified co-activator site called binding function-3 (BF3). This dissertation describes the discovery and development of novel anti-androgens directed towards the BF3 surface of the AR. These inhibitors were developed through a series of computational experiments followed by extensive biological validations. Based on the activity profile of the identified inhibitors, it can be anticipated that these drug prototypes will lay a foundation for the development of alternative or supplementary small-molecule therapies capable of combating PCa even in its drug resistant forms. Because the emergence of castration resistance is the lethal end stage of the disease, we anticipate that the thesis work will eventually have a substantial impact on the survival of prostate cancer patients.

Preface

1. Work described in Chapter 3 has been published [**Munuganti RS**, Leblanc E, Axerio-Cilies P, Labriere C, Frewin K, Singh K, Hassona MD, Lack NA, Li H, Ban F, Tomlinson Guns E, Young R, Rennie PS, Cherkasov A. Targeting the binding function 3 (BF3) site of the androgen receptor through virtual screening. 2. development of 2-((2-phenoxyethyl) thio)-1H-benzimidazole derivatives. J. Med. Chem. (2013) 56, 1136-1148]. Drs. Cherkasov and Rennie are senior authors, and they supervised this project throughout the concept formation to manuscript revision. I performed all the computational experiments, as well as drafted and revised the manuscript. Drs. Leblanc E and Lack NA, Ms. Frewin K and Singh K, contributed to the biological evaluation of the inhibitors. Drs. Guns, Hassona and Mr. Adomat contributed the *in vivo* data. Drs. Labriere C and Young R synthesized the compounds while Mr. Axerio-Cilies and Drs. Ban F and Li H helped with the computations.

2. Work described in section Chapter 4 (except sections 4.2.4 and 4.2.5) has been published [**Munuganti RS**, Hassona MDH, Leblanc E, Frewin K, Singh K, Ma D, Ban F, Hsing M, Adomat H, Lallous N, Andre C, Jonadass JP, Zoubeidi A, Young RN, Guns ET, Rennie PS, Cherkasov A. Identification of a Potent Antiandrogen that Targets the BF3 Site of the Androgen Receptor and Inhibits Enzalutamide-Resistant Prostate Cancer. Chem. Biol. (2014), 3, 1476-1485]. Drs. Cherkasov and Rennie are senior authors, and they supervised this project throughout the concept formation to manuscript revision. I performed all the computational experiments, as well as drafted and revised the manuscript. Drs. Leblanc E, Frewin K, Singh K, Ma D and Lallous N contributed to the biological evaluation of the inhibitors. Drs. Guns, Hassona MDH and Mr. Adomat contributed the *in vivo* data. Drs. Andre C, Jonadass JP and Young R synthesized the compounds. Drs. Ban F and Hsing M helped with the computations while Zoubeidi A has provided Enzalutamide resistant cell line.

Materials and methods described in Chapter 2 (except sections 2.1.6 to 2.1.12) have been published in above mentioned research papers.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABS. androgen binding site

ADME. absorption, distribution, metabolism, excretion

ADT. androgen deprivation therapy

AF1. activation function 1

AF2. activation function 2

AR. androgen receptor

ARE. androgen response element

APDD. atom pair distance dependent

BF3. the binding function 3

CADD. computer-aided drug discovery

CRPC. castration-resistant prostate cancer.

DBD. DNA-binding domain

5α-DHT. 5α-dihydrotestosterone

eGFP. enhanced green fluorescent protein

ER. estrogen receptor

ERG. ETS-related gene

FKBP52. 52 kDa FK506 binding protein

GPU. graphics processing unit

GR. glucocorticoid receptor

HTS. high throughput screening

IGF. insulin-like growth factor

IL. interleukin

LBD. ligand-binding domain

LBDD. ligand-based drug discovery

MD. molecular dynamics.

MR. mineralocorticoid receptor.

NLS. nuclear localization signal.

NMR. nuclear magnetic resonance.

NTD. the N-terminal domain

PCa. prostate cancer.

PTEN. phosphatase and tensin homolog.

PI3K. phosphatidylinositol 3-kinase.

PIP2. phosphatidylinositol 4, 5-triphosphate.

PIP3. phosphatidylinositol 3, 4, 5-triphosphate.

PK. pharmacokinetics.

PPARγ. peroxisome proliferator-activated receptor-gamma.

PR. progesterone receptor.

PSA. prostate-specific antigen.

QSAR. quantitative structure-activity relationship.

R1881. a synthetic androgen.

RMSD. root mean square deviation.

ROC. receiver operating characteristic

SBDD. structure-based drug discovery.

SRC. Steroidal receptor coactivator

TAU. transcriptional activation units

TMPRSS2. transmembrane protease, serine 2.

VS. virtual screening.

Acknowledgements

This dissertation features the results of my research in the course of my PhD study, and it wouldn't be completed without help and support from many talented and inspiring people. First and foremost, I would like to thank my supervisor Dr. Artem Cherkasov for offering me the opportunity to work in his lab at the Vancouver Prostate Centre, which allows me to continue my pursuit in the field of drug discovery. His sharp vision, efficient workstyle, enthusiasm and commitment to work have greatly influenced me and motivated me to keep pushing myself to higher levels. I would like to thank my supervisory committee Drs. Amina Zoubeidi, Colin Collins and Wyeth Wasserman for their mentorship in each stage of my PhD tenure. I am grateful to Dr. Paul S Rennie for his insightful guidance, useful suggestions and strong support in the past five years both as a mentor and leader of our AR inhibitor group.

I would like to thank all members in Drs. Cherkasov, Rennie and Guns labs especially Drs. Eric Leblanc and Takeshi Yamazaki for their help, support and collaboration. I appreciate the experience to work in a multidisciplinary environment, to communicate and learn from people with different backgrounds and expertise. I would like to acknowledge the UBC Experimental Medicine program Director Dr. Vincent Duronio for his help and support throughout my PhD. I also would like to thank Prostate Cancer Foundation, BC for providing financial support during my study.

Words cannot express gratefulness to my parents, brothers and sister for their unconditional love and support that helped me to go from India to Canada to pursue my doctoral study. Without their love and encouragement, it wouldn't be possible for me to complete my PhD degree. I appreciate all the help, encouragement, and support from my dear wife. She shared my joy, complaints, pressures and helped me to stay on track.

Chapter 1: INTRODUCTION

1.1 Prostate Cancer

In recent decades, prostate cancer (PCa) has emerged as one of the most challenging oncologic problems in medicine and the second leading cause of cancer related death in elderly males particularly in western societies.[1] It has been estimated that on an average 24,000 Canadian men will be diagnosed with PCa (24% of all new cancer cases) and 5,000 will die from it every year.[2] PCa is a malignant tumour that starts in cells of the prostate, a walnut-sized gland which is a part of the urinary and reproductive systems of the male. The main function of the prostate is to help control the flow of urine and production of seminal fluid.

PCa is a heterogeneous disease, the etiology of which appears to be related to a complex range of factors including age, race, lifestyle patterns, genetic and nutritional factors.[3] Several studies evidenced that in most cases some form of inappropriate activation of androgen receptor (AR) is linked to recurrent growth of prostate cancers.[4, 5] AR is expressed throughout prostate cancer progression and persists in the majority of patients with hormone refractory disease.[6-8] Also, high levels of AR have been observed in over 80% of locally advanced castration-resistant prostate cancers.[4] Since AR is central to progression to castration resistance, inhibition of this protein remains an important therapeutic approach.

1.1.1 Stages of Prostate Cancer

Staging is a way of describing where the cancer is located and whether it is affecting other parts of the body. There are two types of staging for PCa.

- The clinical stage is based on the results of tests done before surgery, which includes digital rectal examination, biopsy and sometimes X-rays, CT and/or MRI scans, and bone scans. The tests are recommended based on the level of serum prostate specific antigen (PSA), the size of the cancer, which includes its grade and volume.
- The pathologic stage is based on information found during surgery, plus the laboratory results, referred to as pathology, of the prostate tissue removed during surgery. The surgery often includes the removal of the entire prostate and some lymph nodes.

Usually, the urologists use the TNM system to describe the stages of PCa. TNM is an abbreviation for tumor, node and metastasis.[9]

1.1.2 Prognostic Factors

In addition to stage, urologists use other prognostic factors to help plan the best treatment and predict how successful treatment will be (table 1.1). Below are prognostic factors for men with different stages of PCa.

PSA test. PSA is a measurement of *prostate-specific antigen* levels in a man's blood.[10] These results are usually reported as nanograms per milliliter (ng/mL). In some cases, the tumors do not cause an increased PSA level, so a normal PSA does not always mean that there is no PCa.

Gleason score. The Gleason Scoring System is the most common grading system used in PCa. The pathologist looks at how the cancer cells are arranged in the prostate and assigns a score on a scale of 1 to 5. Cancer cells that look similar to healthy cells are given a low score, and cancer cells that look less like healthy cells or are more aggressive looking are given a higher score. To assign the numbers, the doctor determines the main pattern of cell growth, which is the area where the cancer is most obvious, looks for any other less common pattern of growth, and gives each one a score. The scores are added to come up with an overall score between 2 and 10. For example, a Gleason score of 7 is a medium-grade cancer, and a score of 8, 9, or 10 is a high-grade cancer. [11]

Table 1.1. Clinical stages of prostate cancer.

Stage	Test results	Clinical stage	Description	Treatment
1	PSA <10ng/mL and Gleason score <=6	T1a, N0, M0	Cancer growth is slow and no symptoms	Radical prostatectomy, Radiation
2	PSA <10ng/mL and Gleason score <=6	T1b-T2, N0, M0	Cancers have not grown outside the prostate but are larger than stage 1	Radical prostatectomy, Radiation
3	PSA >=10 but <20ng/mL and Gleason score =6	T3, N0, M0	Cancers have spread beyond the prostate but have not reached other organs	External beam radiation + hormone therapy, hormone therapy alone
4	PSA >=20ng/mL and Gleason score >=8	T4, N1, M1	Cancers have spread to the bladder, rectum, lymph nodes and bones	External beam radiation + hormone therapy, hormone therapy alone

1.1.3 Castrate-Resistant Prostate Cancer (CRPC)

Although radiation therapy and radical prostatectomy are used to treat PCa (depending on the stage of disease, advice of the physician and choice of the patient), androgen deprivation therapy (ADT) is considered to be the 'gold standard' treatment option for PCa.[12] As the name suggests ADT blocks either the production or the function of androgens. ADT involves treatment with luteinizing hormone-releasing hormone antagonists and AR inhibitors.[13] While treatment with AR antagonists can initially suppress the prostate cancer growth, long-term hormone therapy becomes progressively less effective. This leads to the progression of surviving tumor cells to the castration resistant state.[14] CRPC presents a spectrum of diseases ranging from rising PSA levels without metastases to bone metastases. [15, 16] It has been reported that bone metastases is common phenomena, affecting approximately 90% of men with CRPC, causing severe pain and spinal cord compression.[17, 18]

Although the progression to CRPC is not fully understood, studies have determined several different mechanisms involving AR that eventually lead to CRPC.[19, 20] Briefly, there are four critical mechanisms for the development of castration resistance. 1) Overexpression of AR 2) AR mutations that render the receptor responsive to non-androgen ligands including antagonists 3) Expression of AR splice variants 4) Androgen independent AR activation mechanisms. These factors will be discussed in detail in the following chapters.

1.2 Androgens and Androgen Receptor

Androgen receptor can be activated by the binding of its physiological ligands, testosterone and 5α-dihydrotestosterone (5α-DHT). [21, 22] Physiologically, functional AR is responsible for in utero male sexual differentiation and for male pubertal changes. In adult males, androgen is mainly responsible for development and the maintenance of male secondary characteristics including spermatogenesis and bone mineral density.[22] The function of androgens can be classified into 1) androgenic effect-occurs in the reproductive tissues, including prostate and testis 2) anabolic effect-occurs in muscle and bone.[23]

The human AR, classified as NR3C4 (nuclear receptor subfamily 3, group C, gene 4) belongs to the steroid hormone group of nuclear receptor superfamily. AR shares high sequence homology with other members of the family that includes estrogen receptor (ER), glucocorticoid receptor (GR), progesterone receptor (PR) and mineralocorticoid receptor (MR). [24, 25] AR is a

ligand-dependent transcription factor mainly expressed in androgen target tissues, with the highest expression level observed in the prostate, adrenal gland and epididymis. [26] Therefore, AR is associated with many androgen-regulated clinical disciplines ranging from urology (prostate cancer), neurology (spinal bulbar muscular atrophy) to dermatology (hirsutism, baldness and acne).[27]

Briefly, synthesis of testosterone occurs primarily by the Leydig cells in the testes, under the regulation of luteinizing hormone, which is in turn regulated by gonadotropin-releasing hormone. Once produced, testosterone mostly circulates bound to serum sex hormone-binding globulin [28] and albumin.[29] Free form of testosterone enters prostate cell and gets converted into a more potent metabolite DHT by the action of 5α-reductase enzyme.

1.2.1 Activation of AR

The activation of AR follows a well characterized pathway. In the cytoplasm of prostate cells, unliganded AR associates with a complex of cytoplasmic factors and molecular chaperone heat shock proteins 40, 70 and 90 that maintain the structural integrity of the receptor in a high-affinity agonist binding conformation.[30] Upon binding of DHT, the AR undergoes a series of conformational changes, including disassociation from chaperones, interactions between its N and C termini, interaction with co-factors such as importin-α, which transports proteins across the nuclear pore complex into the nucleus, dimerization and binds to specific androgen response element (ARE).[31, 32] After translocation, AR interacts with co-activator proteins such as steroid receptor coactivators (SRCs) at the co-activator binding site called activation function (AF2) site.[33] This triggers the recruitment of RNA polymerase II and other transcriptional factors (p300, CREB binding protein etc.) to form a complex with the AR that leads to the transcription of target genes promoting the growth and survival of prostate cells (figure 1.1). Previously, it was hypothesized that the nuclear AR is phosphorylated by kinases, forms a dimer and binds to AREs found in the promoters and enhancers of AR-dependent genes. However, in 2012 van Royen *et al* suggested that dimerization of the AR only occurs after nuclear translocation and may require prior binding to the DNA.[34]

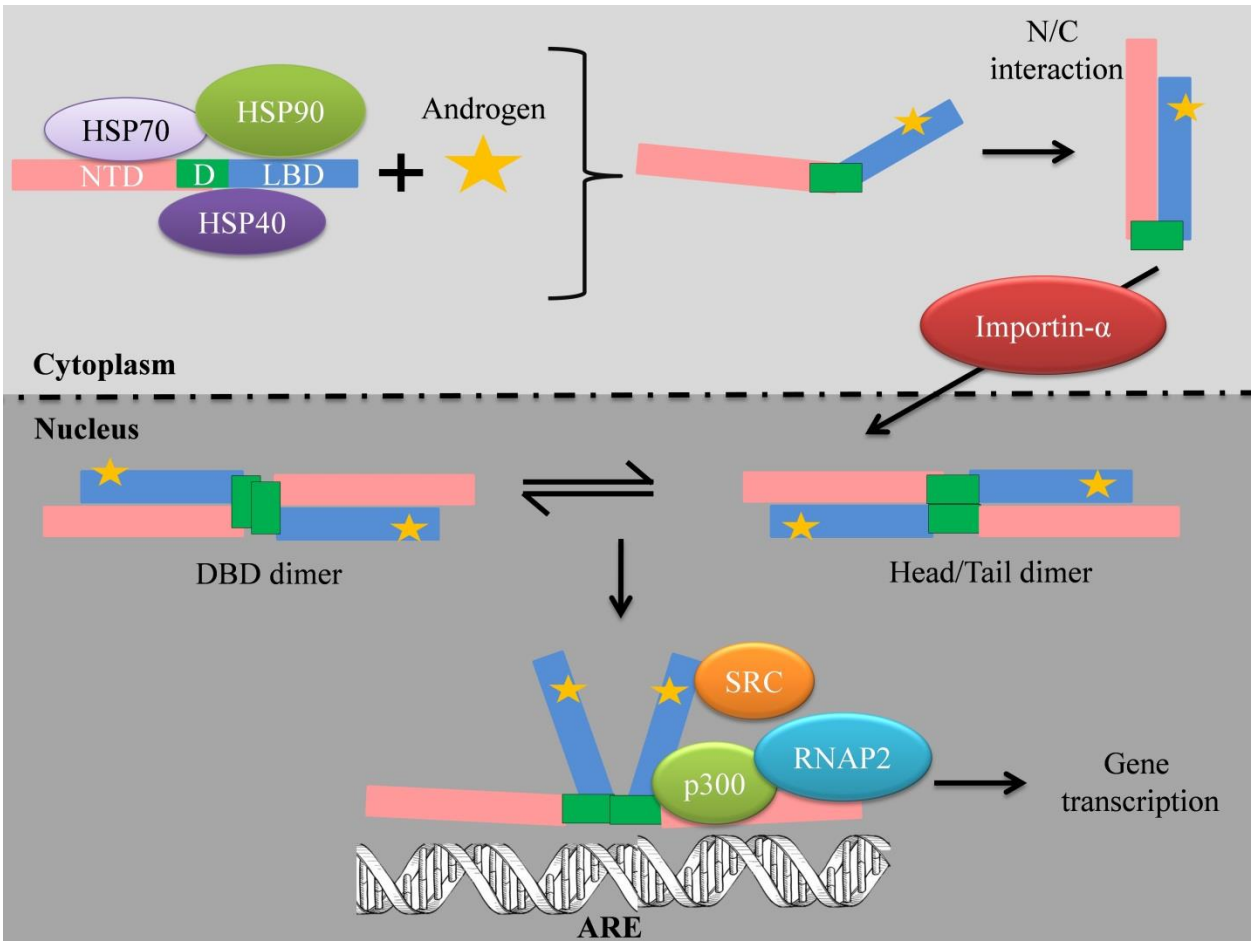


Figure 1.1. The signaling pathway of androgen receptor. The AR domains are labeled as. NTD - N-terminal domain, D - DNA binding domain, LBD - ligand binding domain.

1.2.2 Structural and Functional Aspects of AR

The AR gene (located on the X chromosome at Xq11-12) is more than 90 kb long and codes for a protein of 919 amino acids. Even though steroidal nuclear receptors are implicated in different physiological processes, they all share the same modular structure and domain organization.[35] In AR, the N-terminal domain (NTD) is encoded by exon 1. Exons 2 and 3 encode for the DNA-binding domain (DBD) whereas the ligand binding domain (LBD) is encoded by exons 4 to 8. There is also a small hinge region between the DNA-binding domain and ligand-binding domain (figure 1.2A).[21]

N-Terminal Domain (NTD). This region accounts for more than 60% (1 to 558 residues) of the AR protein and functions as a potent transcriptional activator independent of the androgen activation. Based on sequence similarity, the NTD is the least conserved domain amongst all nuclear receptors. Human AR shares only 8.4% sequence identity with the NTD of

human ER, 14.9% identity in the case of the GR and 21.9% with the PR. Structurally, the AR NTD is highly flexible and displays intrinsic disorder in solution due to which elucidation of its crystal structure is almost impossible. [36-38] However, biophysical study of the NTD has revealed that it exists in a molten globule conformation[38] with regions of rigid secondary structures. It was hypothesized that these regions are either buried or exposed in response to different cellular events including levels of androgen, type of coactivators and chromatin environment.

The NTD is characterized by the presence of two transcriptional activation units (TAU) termed TAU-1 (residues 101–370) and TAU-5 (residues 360–485). [39, 40] Importantly, the AR-NTD contains two motifs that are essential for the interaction with its ligand binding domain *i.e.* FxxLF motif (at residues 23–27) and WxxLF motif (at residues 433–437). The corresponding contact (termed N/C interaction) has been shown to be critical for stabilizing the androgen in the ligand binding pocket and for overall AR function.[41, 42] Furthermore, the NTD contains two large ploy-amino acid repeats known as homopolymers *i.e.* poly-glutamine and poly-glycine fragments, averaging 21 and 24 residues, respectively. Several studies have been reported that any variation in length of the poly-glutamine can result in Kennedy disease (also known as X-linked spinal and bulbar muscular atrophy) and prostate cancer. [43, 44] Deletion of portions of the NTD or mutation of some of its key residues such as I229A, M244A, L246A, and V248A causes a decrease of AR transactivation activity. [45]

The AR NTD plays a critical role in mediating AR transcriptional activity by recruiting several transcription machinery components including the members of Transcription factor II, [46] coactivator proteins such as CREB-binding protein [47, 48] and co-repressors, like SMRT.[49]

The DNA Binding Domain (DBD). This domain contains 66 amino acids with two zinc finger motifs, where each metal ion is coordinated by four cysteine residues. The residues GSCKV (577-581) forms the P box that interacts with the major groove of the DNA, while the second zinc finger contains the D box (residues 596–600. ASRND), which plays a role in DBD-mediated AR dimerization (Figure 2B). [50] The AR DBD recognizes classical AREs on the DNA that are organized as inverted repeats of 5'-AGAACA-3'-like motifs with a three nucleotide spacer and selective AREs that are considered direct repeats of 5'-AGAACA-3'-like motifs.[51, 52]

The DBD is highly conserved domain and shares high degree of sequence similarity with other nuclear receptors (79.5% identity with PR, 71.2% with GR and 53.4% with ER).[21] Shaffer *et al* solved the rat AR DBD in a complex with a direct repeat of ARE (PDB. 1R4I) and reported

that the AR DBD is formed by two short anti-parallel β-strands and two perpendicular α-helices (figure 1.2B). The study concluded that this particular structural organization allows the AR DBD to bind to the DNA in the form of a "head to head" dimer, where one monomer binds the half-site response element with high affinity and the second binds the other half-site with lower affinity.[53]

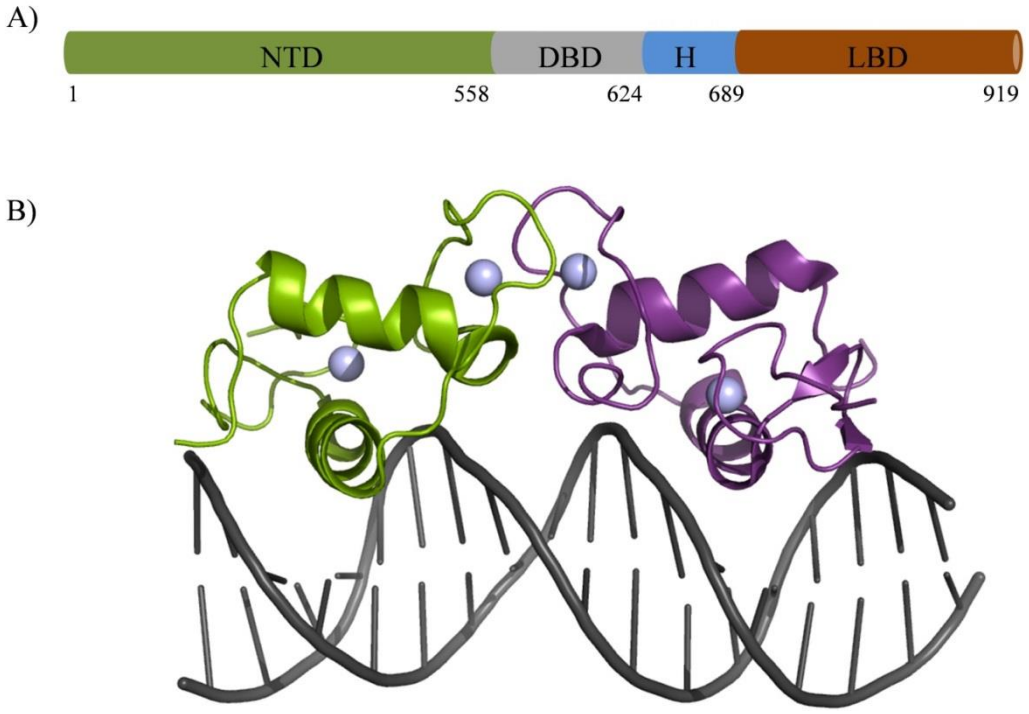


Figure 1.2. Structural details of AR DBD region. A) Domain organization of the androgen receptor. B) Cartoon representations of the rat AR-DBD structure. Zinc ions are presented as spheres and the D-box is highlighted in green. The P-box is shown as purple color.

Hinge Region. Although this region (residues 625–689) is flexible and poorly conserved among nuclear receptors, diverse functions have been ascribed to it. AR hinge region mediates its transcriptional activity by acting as a major site for posttranslational modifications including methylation[54] and ubiquitylation.[55] The signal responsible for nuclear import is encoded by a bipartite nuclear localization signal (617-RKCYEAGMTLGARKLKKL-634) formed by two clusters of basic residues belonging to the C-terminus of the DBD and the N-terminus of the hinge region.[56] The nuclear translocation of AR occurs in a step-wise manner. First, when androgen binds to unliganded AR in cytoplasm, it undergoes a conformational change, which exposes the nuclear localization signal and facilitates its interaction with importin-α–importin-β

complex. This association results in translocation of the activated AR to the nucleus. Crystallographic analysis of importin-α-AR complex revealed that residues 629-RKLKKL-634 are critical for binding, with an important contribution from Lys630 (figure 1.3). [31]

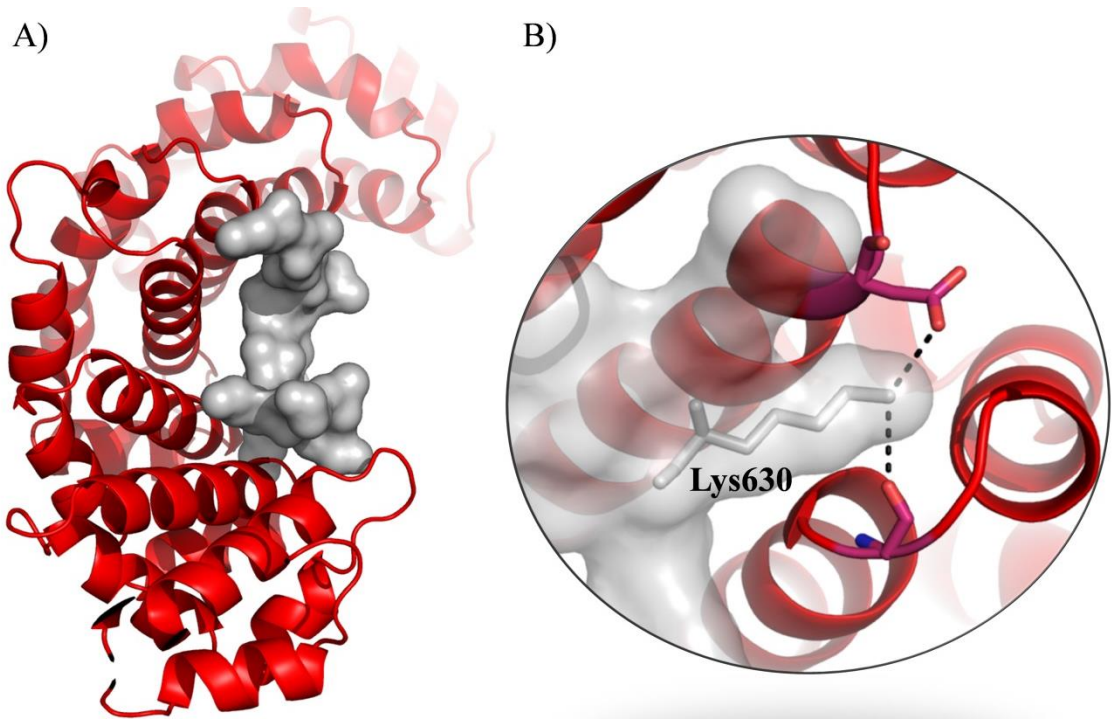


Figure 1.3. Structural details of AR hinge region. A) Cartoon representations of the importin α and β in complex with AR nuclear localization sequence (grey surface). B) AR residue Lys630 forms hydrogen bond interactions with importin. These interactions are shown as black dotted lines.

Ligand Binding Domain (LBD). AR LBD (in fact, all steroid receptor LBDs) performs a conventional mechanistic function *i.e.* binding its physiological ligands at hormone binding sites and inducing structural changes which ultimately forms the coactivator binding surfaces (figure 1.4A).[57] The role of the AR LBD is of particular importance for PCa, because it is the foremost target of current androgen deprivation therapies.

Several crystal structures of AR LBD have been solved by X-ray crystallography, revealing that it is composed of 11 α-helices (numbered 1–12, where helix 2 is missing compared to other NRs), arranged as a three-layered helical sandwich and four β-strands organized in two short sheets. Figures 1.4B shows that a total of 18 residues belonging to β1 and helices 3, 5, 7 and 10 forms androgen binding site (ABS) and make hydrogen bonds and/or hydrophobic

interactions with androgens (figure 1.4C).[58] Upon agonist binding, helix 12 (H12) is repositioned and serves as the "lid" of the ligand-binding pocket to stabilize the ligand, and the very end of the C-terminal region of the LBD forms the second α-turn (next to H8 and H10), which works as a "lock" to further stabilize the "lid" (H12) conformation.[59] The agonist-induced conformational change in the LBD allows the formation of AF2 pocket on the surface of LBD (green colored region in figures 1.4A and D), which is crucial for both the N/C interaction of AR and co-regulator recruitment during transcriptional activation.[21]

AF2 pocket is a hydrophobic groove on the AR surface, which is flanked with regions of positive and negative charges, "charge clamps", that are essential for binding AR activation factors. AF2 is highly conserved a protein interaction site and has also been extensively analyzed throughout this nuclear receptor family. It recruits numerous coactivators, including members of the p160 SRC family, such as SRC1, 2 and 3.[33] Crystal structures of AR complexed with SRC2 and SRC3 have revealed that Lys720 and Glu837 act as charge-clamp residues and stabilize the AF2-coactivator interactions (Figure 1.4D).[60] Importantly, the AR AF-2 domain displays a higher affinity for NTD-derived FxxLF-containing peptides than coactivator-derived LxxLL-containing peptides, suggesting that N/C interaction, rather than direct transcriptional activation, may be the primary role for AF2.[61] Moreover, AF2 preferentially binds the NTD when the AR is mobile, but also recruits co-regulators when AR engages with DNA.[62] Therefore, these findings suggest that N/C interaction may block inappropriate co-regulator interaction until the AR is engaged with AREs in the promoter and enhancer regions of target genes. After AR is bound to DNA, the AF2 cleft may be more amenable to coactivator binding, which would enhance the transcriptional activity of the AR.

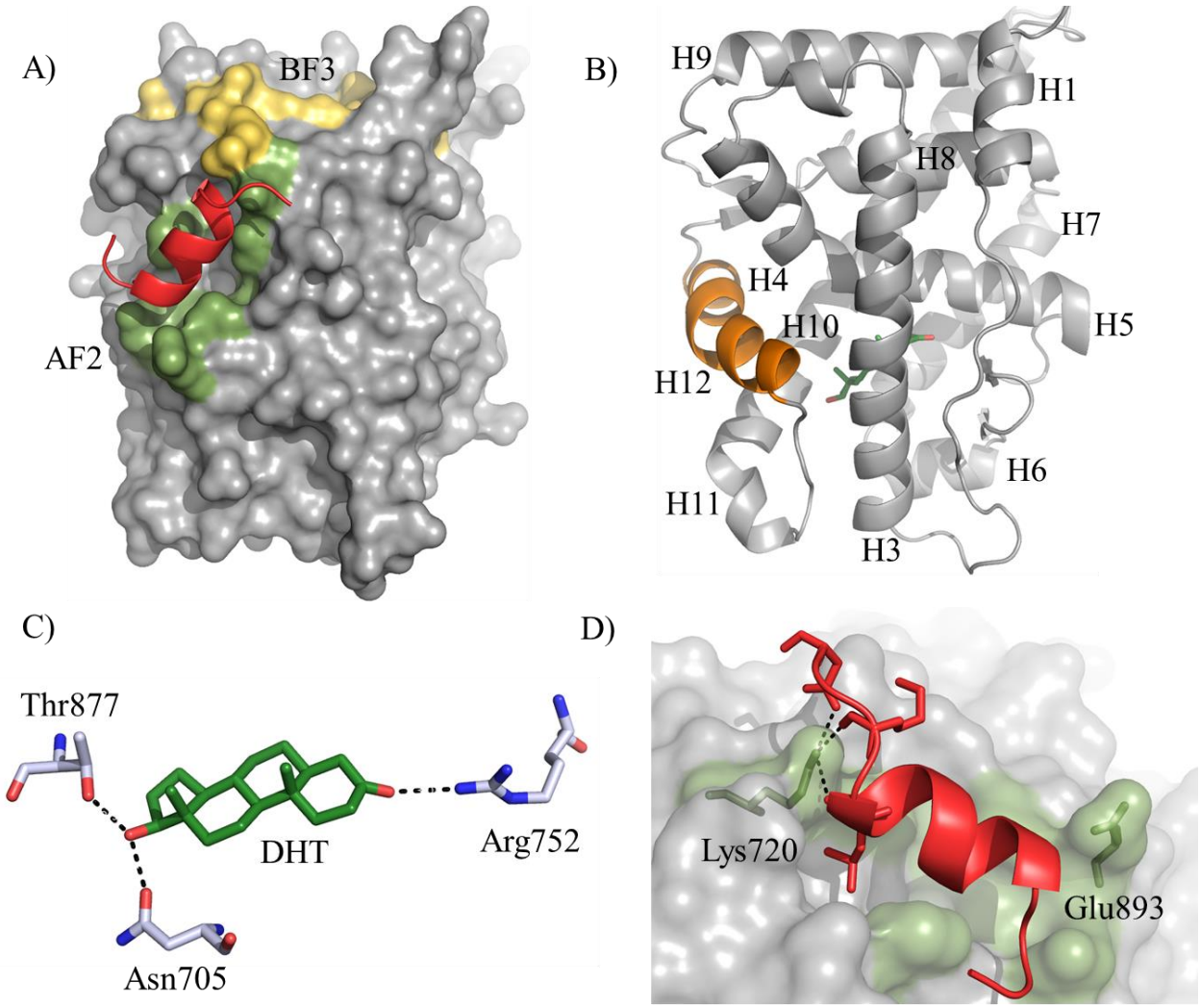


Figure 1.4. Structural details of AR LBD region. A) Surface representations of the AR ligand binding domain. AF2 pocket is shown green color whereas BF3 site is shown in yellow color. SRC peptide is represented by a red helix. B) Cartoon representations of the AR ligand binding domain. Helix 12, which plays a critical role in formation of the AF2 pocket, is highlighted in orange color. DHT is shown as green stick. C) A network of hydrogen bond interactions between DHT and AR ABS residues. These interactions are critical for the tight binding of the native ligand in the cavity. D) Charge clamp residues at AF2 pocket, Lys720 and Glu893, which stabilize the interaction with LxxLL motif, are highlighted in green color. SRC peptide is shown in red color.

Several lines of evidence demonstrate that the AF2 groove may not be the sole protein–protein interface dictating macromolecular assembly upon LBD engagement in various nuclear receptors. Using a combination of X-ray crystallography and functional assays, Estébanez-

Perpina *et al* identified a novel site on the AR LBD surface and named it binding function-3 (BF3).[63] The BF3 pocket is located adjacent to the AF2 site but distant from the ABS (figure 1.4A). As it can be observed from the X-ray structures, the BF3 pocket is formed by residues from several LBD-forming helices. The residues contributing to BF3 formation are. Pro671, Ile672, and Phe673 from the NH2-terminal part of a helix 1 (H1), Pro723, Gly724, Arg726, and Asn727 from H3, and Phe826, Glu829, Leu830, Asn833, Glu837, and Arg840 from H9. The residues Arg726 and Asn727 act as boundary between the AF2 and BF3 sites and may play an important role in their cross talk and coordinated action.[63, 64]

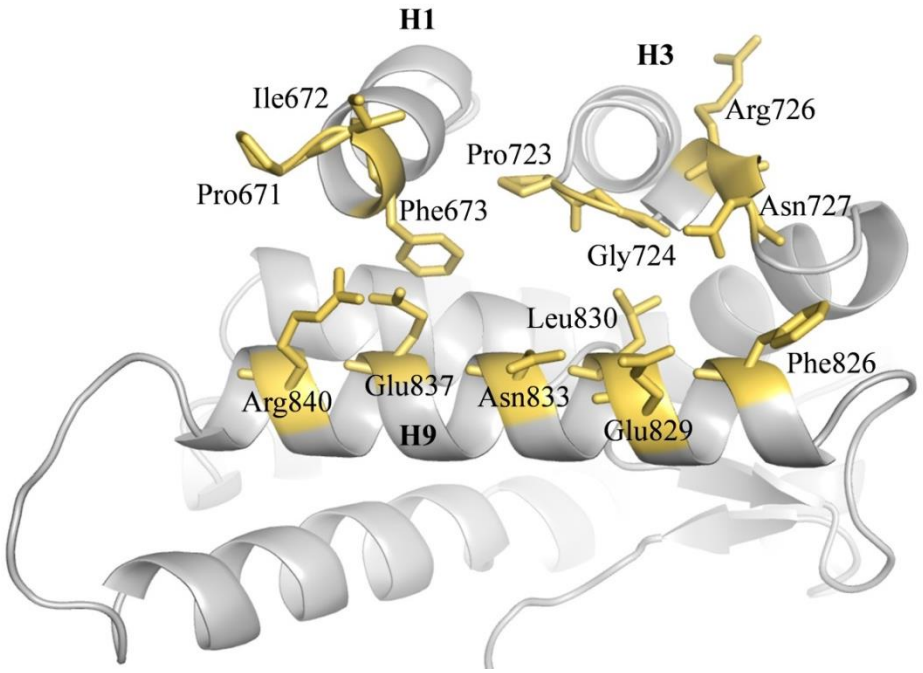


Figure 1.5. Cartoon representations of the BF3 site on the surface of AR LBD. Residues of helix 1, 3 and 9 form this pocket. Critical residues are shown in yellow.

A 52 kDa protein called FK506 binding protein (FKBP52) which is known to be an important positive regulator of AR has been shown to function through the BF3 domain.[65] FKBP52 is a cochaperone of Hsp90 that binds the AR/Hsp90 complex and regulates ligand binding to the AR. Recently, Jehle *et al* showed that N-terminal amino acid sequences of Bag-1L bind to the AR and contribute to the transactivation function of the receptor. Furthermore, they demonstrated that a hexapeptide motif "GARRPR" in Bag-1L binds to BF3 pocket.[66] Single amino acid exchanges in the BF3 pocket destroy binding of the Bag-1L peptides to the LBD in a

mammalian two-hybrid assay. It has been documented that in addition to known mutations in the ABS, the BF3 area is also associated with PCa and androgen insensitivity syndromes.[67] Given the importance of this region for AR function and modulation, it clearly represent prospective targets for developing novel PCa therapeutics.

1.3 Current Advances in the Development of AR Inhibitors

Since androgen receptor is considered as an important therapeutic target for PCa and CRPC, number of inhibitors have been reported in the literature with different mechanisms of action. Based on their functionality and target domain, they can be categorized into AR ABS inhibitors, Selective Androgen Receptor Modulators (SARMS), AR DBD inhibitors and AR NTD inhibitors.

1.3.1 AR ABS Inhibitors

This class of inhibitors competes with cognate ligands and prevents their binding to the receptor. They can be classified as steroidal and non-steroidal based on their chemical scaffold. The first anti-androgen to be tested in the clinic was Cyproterone acetate, back in the early seventies[68] (figure 1.6). It is a weak anti-androgen even though it binds to AR with relatively high affinity compared to other anti-androgens. Other known steroidal anti-androgens are Oxendolone and Spironolactone (figure 1.6). However, they demonstrated several clinical limitations including potential hepatotoxicity and cross reactivity with other nuclear receptors. More importantly, due to steroidal skeleton, structural modification of these compounds was restricted. Therefore, steroidal anti-androgens are rarely used in the clinic.

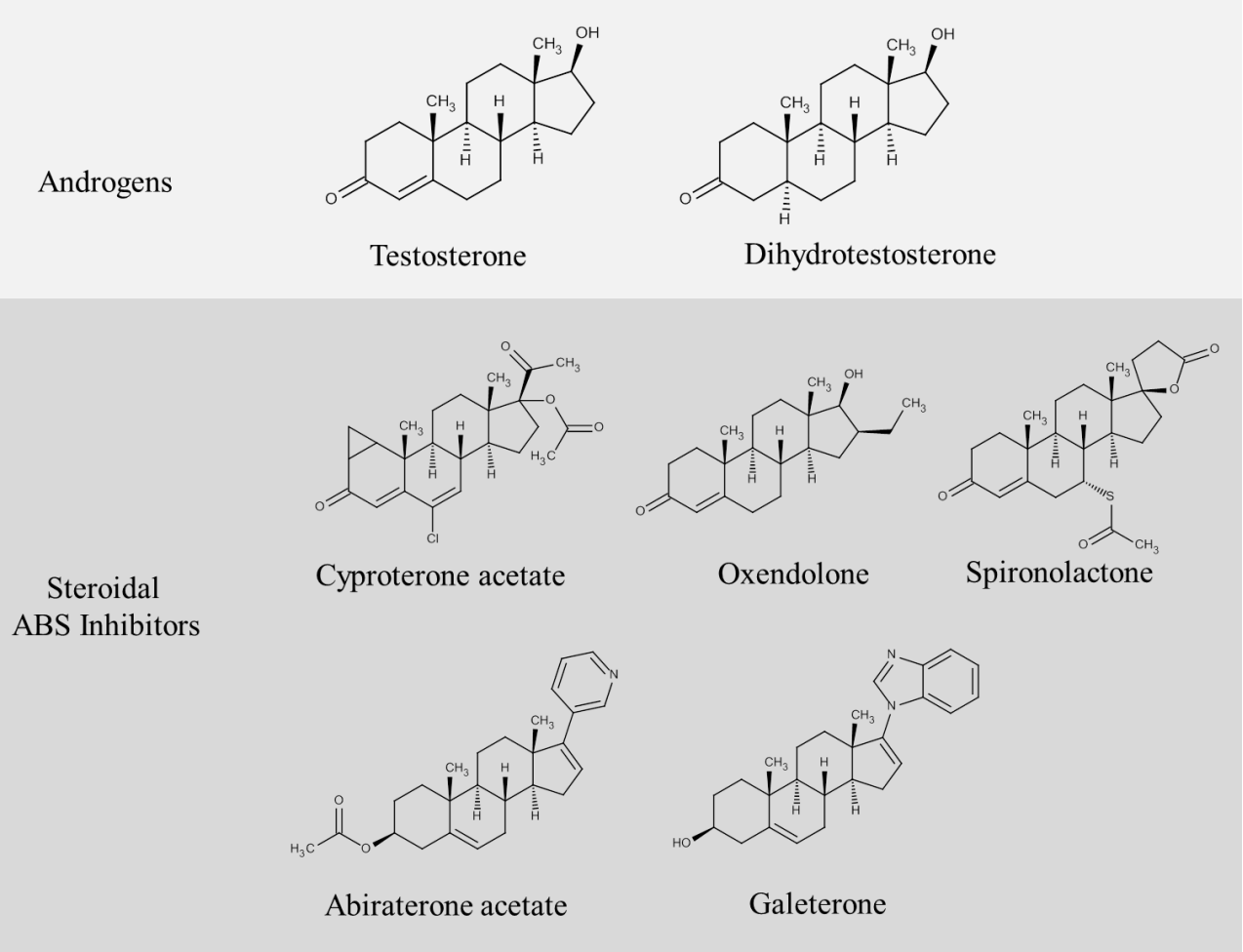


Figure 1.6. Chemical structures of androgens and steroidal anti-androgens.

Subsequently, non-steroidal anti-androgens were developed to address the limitations of steroidal inhibitors.[69] Non-steroidal AR ligands are broadly classified into toluidides and hydantoins (figure 1.7). The toluidide derivatives are the first reported non-steroidal compounds. The first generation of anti-AR PCa drugs such as Flutamide (and its derivative, hydroxyflutamide), Nilutamide and Bicalutamide belong to this chemical class. Unlike steroidal compounds they do not possess any significant intrinsic androgenic activity and cross reactivity with other nuclear receptors. Although Flutamide and Nilutamide were initially used as PCa therapeutics, they were replaced by Bicalutamide since it demonstrated better pharmacokinetic profile *i.e.* less hepatotoxicity and longer half-life.[70]

The second generation anti-androgens belong to diarylthiohydantoin chemical class. It includes recently FDA approved drug Enzalutamide and its derivative, investigational drug, ARN509. Similar to toluidide derivatives, Enzalutamide and ARN509 bind to the AR ABS but

with greater affinity (over 5-10 times more potent than bicalutamide).[71, 72] These drugs not only demonstrate improved anti-AR potency, PK profile and enhanced *in vivo* efficacy but also prevent AR nuclear translocation and impair DNA binding.

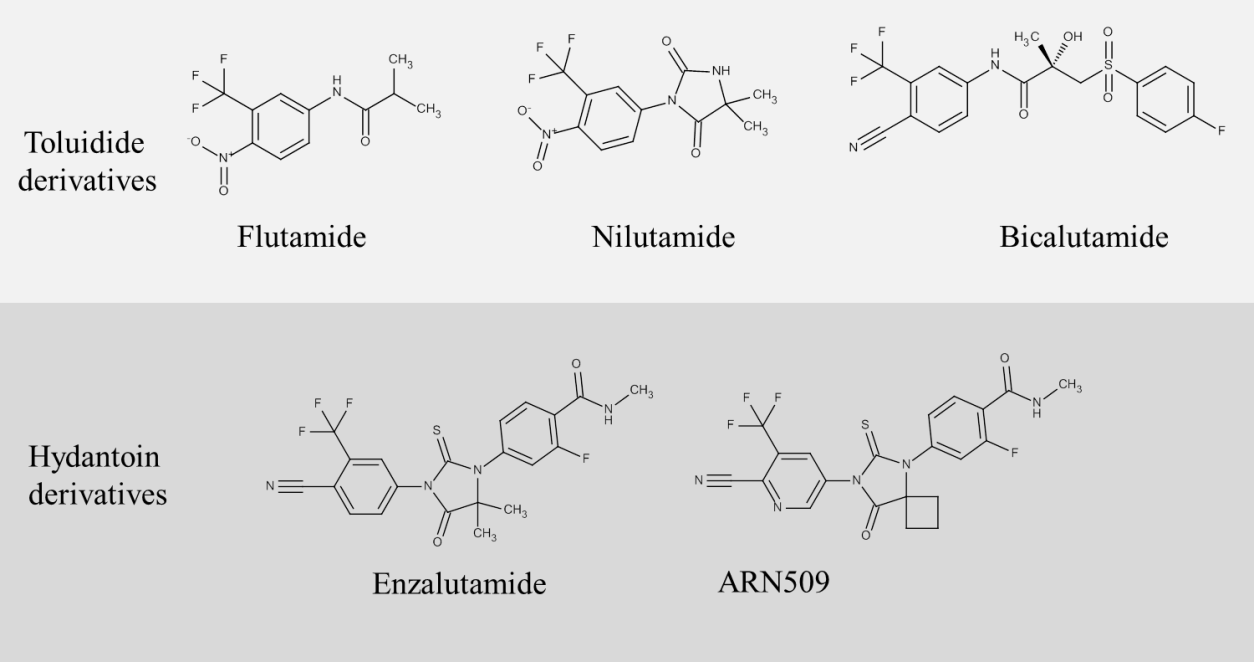


Figure 1.7. Chemical structures of commercially available and experimental anti-androgens that target androgen binding site of AR.

Other anti-androgens known for PCa therapy are Abiraterone acetate and Galeterone/TOK-001 (figure 1.5). Abiraterone acetate, a prodrug of abiraterone, is a selective inhibitor of androgen biosynthesis that potently blocks cytochrome P450 c17 (CYP17), a critical enzyme in testosterone synthesis, thereby blocking androgen synthesis by the adrenal glands and testes and within the prostate tumour. [73, 74] The anti-AR profile of Abiraterone contributes to its anti-tumoural effects and is clinically used for metastatic CRPC.[75] Galeterone is a proprietary small molecule, oral drug for the treatment of CRPC that disrupts AR signaling. Galeterone selectively inhibits CYP17 lyase to prevent testosterone synthesis, blocks androgen binding to the AR, prevents AR binding to chromatin and degrades the AR protein.[76, 77]

ODM-201 is an investigational anti-androgen that is thought to block the growth of PCa cells. ODM-201 has a great affinity binding to the AR and blocks AR nuclear translocation and unlike other anti-androgens, crosses the blood-brain barrier.[78]

1.3.2 Selective Androgen Receptor Modulators (SARMS)

SARMS are known to exhibit tissue selective profile *i.e.* they act as antagonists in prostate tissue but agonize the receptor in pituitary and muscle cells.[79, 80] SARMS are developed with an aim of having potential of PCa treatment while agonistic activity in the muscle and bone can treat diseases such as muscle-wasting conditions and hypogonadism.[80] Based on the structure, SARMS can be classified as steroidal and non-steroidal.

The steroidal SARMs are formed by modifying the chemical structure of testosterone molecule. Pioneering efforts by scientists at Ligand Pharmaceuticals and the University of Tennessee provided the early foundations of the nonsteroidal SARM discovery.[81, 82] Since then, a number of structural categories of SARM pharmacophores have been explored. The aryl propionamides were the first reported SARMS demonstrating better *in vivo* tissue selectivity. This chemical series is structurally similar to bicalutamide and the replacement of aryl sulfonyl in bicalutamide with phenoxyl group leads to the transformation of the antagonist to agonist, such as investigation drugs Andarine and Enobosam (figure 1.7). Currently, these compounds are in clinical trials for the treatment of cachexia and osteoporosis. [80, 83]

1.3.3 AR DBD Inhibitors

There has been little development of inhibitors that specifically target the DBD portion of the AR. Previously, a hairpin pyrrole-imidazole polyamide was designed to target the ARE to disrupt AR DBD binding and was effective at inhibiting androgen-induced PSA expression in LNCaP cells.[84] A follow up study reported a cyclic polyamide that decreases PSA mRNA levels with a greater affinity and specificity towards ARE. [85] Cherian *et al* conducted a high-throughput screen to target the AR and identified a compound that reduces AR-specific DNA binding, although it remains unclear whether this compound works directly by binding to the DBD.[86]

Given that the structure of the AR DBD has been solved,[53] our group conducted a systematic computational screen and identified small-molecule inhibitors that disrupt AR DBD−DNA interaction.[87, 88] The authors reported two compounds VPC-14228 and VPC-14449 that selectively interact with the intended binding site on DBD and inhibit the growth of Enzalutamide-resistant cells as well as block the transcriptional activity of both full length AR and constitutively active variants. Currently efforts are focused on developing DBD inhibitors with greater potency and stability.

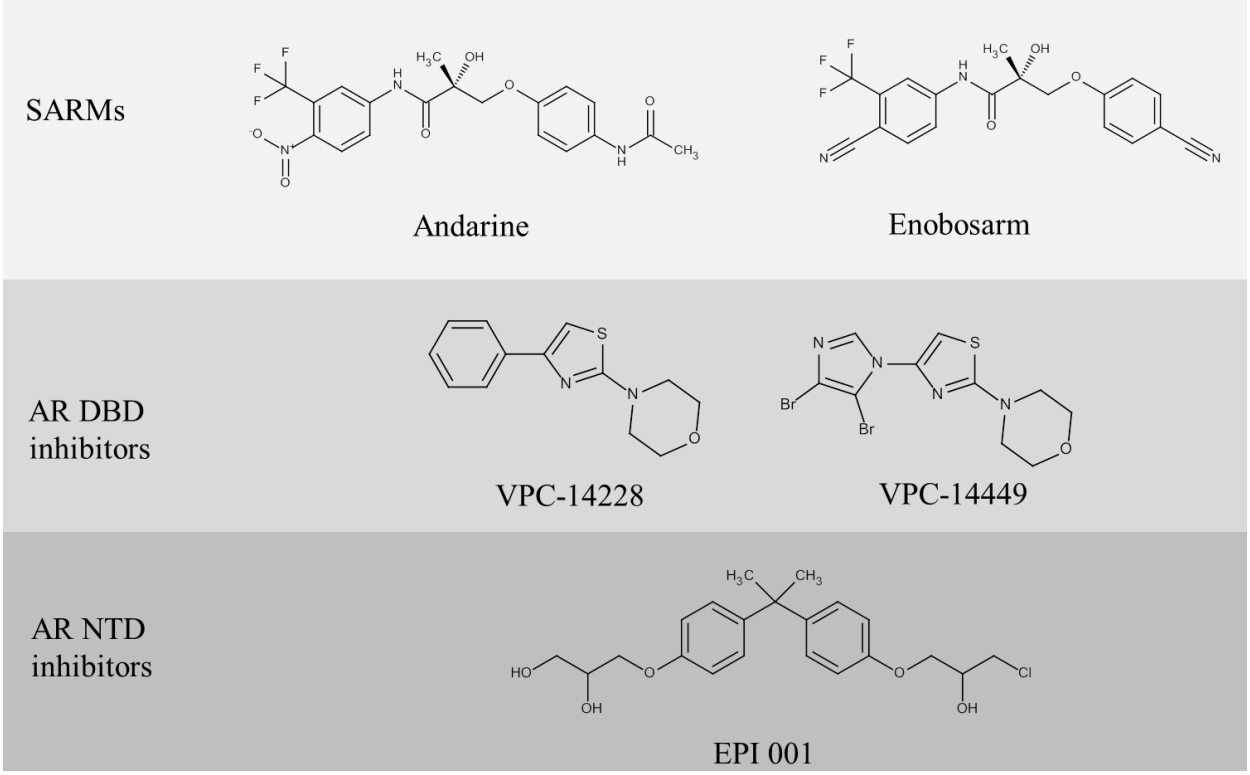


Figure 1.8. Chemical structures of clinically available SARMS and small-molecule inhibitors that target AR DNA binding domain and N-terminal domain reported in the literature.

1.3.4 AR NTD Inhibitors

Since both ligand-dependent and -independent transcriptional activity of the AR is attributed to its N-terminal TAU1 and TAU5 regions, the NTD remains a very attractive drug target for treating both early stage PCa and CRPC. Even though rational drug discovery efforts are hindered due to the unavailability of structural information on NTD, high throughput screening approaches have been so far successful in discovering different classes of NTD inhibitors. EPI-001 was the first small-molecule inhibitor reported to bind covalently to AF1 region and inhibit protein-protein interactions necessary for AR transcriptional activity and its splice variants.[89, 90] Although EPI-001 was identified as a AR NTD inhibitor, it is reclassified as a selective modulator of peroxisome proliferator activated receptor-gamma (PPARγ). [91] Cyclical peptides termed sintokamides[90] and decoy peptides[92] containing the AR-NTD sequence are also reported to inhibit AR NTD.

1.4 Limitations of Commercial Anti-AR PCa Drugs

1.4.1 Overexpression of AR

High AR expression in CRPC at both mRNA and protein levels represents a mechanism of acquired resistance to ADT. Chen *et al* presented an evidence supporting that the AR mRNA is universally upregulated in hormone-refractory xenograft models, and increased AR expression is sufficient to convert Bicalutamide from antagonist to agonist.[93] AR overexpression can occur due to gene amplification[94] and it has been reported that AR gene amplification is present between 20-33% of patients with recurrent PCa who initially responded to ADT for 12 months. [95]AR amplification was not found in untreated primary tumors suggesting it is not involved in the genesis or progression of PCa in patients untreated with ADT.[95] AR amplification promotes the survival and proliferation of tumor cells even if the residual androgen level is low after castration. Since the rate of AR amplification is low, it cannot account for most cases of resistance; in addition, high AR expression at mRNA level without increase in gene copy number has also been observed in CRPC.

1.4.2 Mutations in the ABS

Another well-known mechanism for the development of CRPC is ligand promiscuity of mutated forms of the AR. It has been reported that mutations in AR gene can cause amino acid substitutions in the LBD, more specifically in ABS that hampers the efficacy current drugs (eg, T877A, L701H, W741L, and F876L).[96-98] So far, no crystal structure of the wild-type AR complexed with an antagonist has been reported, due to the instability of the receptor upon antagonist binding. However, structures for W741L-bicalutamide and T877A-cyproterone acetate complexes have been solved.[99, 100] Based on the crystallographic observations it was proposed that AR can accommodate larger atoms due to extra space provided by the T877A mutation. This causes aberrant AR activation in response to hydroxyflutamide, cyproterone acetate and alternative steroids. The AR T877A mutation has historically been of great interest, because it is found in the LNCaP cell line as well as cases of advanced PCa. [101] This mutant thus serves as the prototype for altered AR ligand specificity caused by AR mutations in PCa.

Recently, Balbas *et al* and Korpal *et al* identified the AR mutation F876L that converts Enzalutamide into an agonist.[97, 102] F876L causes the mutant receptor to bind Enzalutamide six times more effectively than wild-type receptor. As the structure of AR-Enzalutamide has not been solved yet, Balbas *et al* modeled the AR-Enzalutamide complex through ligand docking

and molecular dynamics simulations. [97] They suggested that leucine lacks the favorable contact with Enzalutamide that is predicted to be necessary for H12 displacement. Hence, H12 is thought to assume an agonist-like conformation that allows coactivator recruitment.

1.4.3 Splice Variants

There is increasing evidence to show that AR signaling could occur in complete absence of androgen binding, owing to the expression of constitutively active AR variants lacking the LBD portion (figure 1.9).[103, 104] These variants arise primarily through exon skipping and cryptic exon inclusion. Although some splice variants, such as AR-45,[105] are found in normal prostate tissue, variants lacking the LBD have been found to be upregulated in tumours, compared to levels in normal prostate cells.[103, 104, 106] AR-V7/AR3 and AR^{v567es} are the most commonly found variants in PCa and thus the most studied. RNA and protein for both variants has been found in PCa cell lines, xenografts, and human tumor specimens.[107] Recently, Li *et al* demonstrated that expression of constitutively active truncated AR splice variants are the key drivers of AR activation causing resistance to therapies targeting full-length AR, including Enzalutamide.[108]

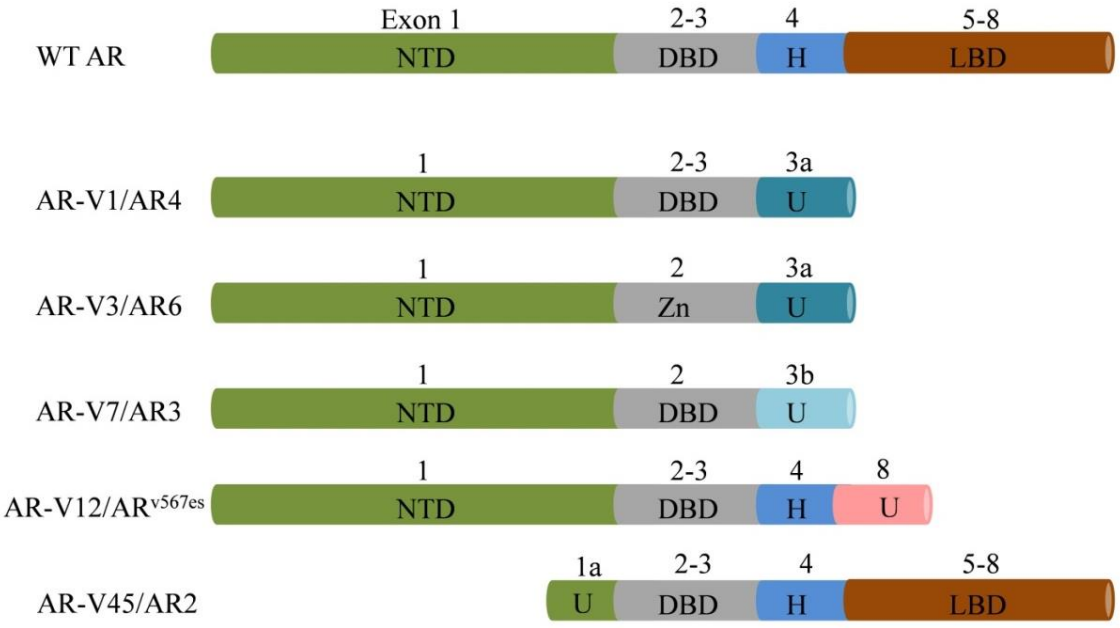


Figure 1.9. Reported splice variants of the AR. The AR gene is organized as eight exons, which form the coding sequence for its different domains. Shown above each splice variant are the corresponding exon numbers included in spliced mRNA. Cryptic exon inclusion results in

unique (U) regions with novel nucleic acid sequences not found in the wild-type AR. In AR-V3/AR6, the inclusion of exon 2 yields a splice variant bearing only one zinc finger (Zn) and, thus, a truncated DBD.

1.4.4 Alternative Pathways

In addition to androgen-involving mechanisms, activation of AR can be mediated by growth factor signaling pathways such as epidermal growth factor receptor,[109] insulin-growth factor-1 [110] and interleukin-6.[111] For example, insulin-growth factor 1 was able to cause AR activation, inducing a five-fold increase in PSA levels in LNCaP cells cultured in serum-free medium. Activation of the AR complex can also occur via crosstalk with other signaling pathways, such as those mediated by the non-receptor tyrosine kinases including Src kinase.[112] Besides, the PI3K/mTOR signaling pathway is a critical oncogenic pathway that plays a role in the tumorogenesis and resistance in a variety of cancers. The pathway was found to be deregulated in advanced PCa and associated with ADT resistance.[113] Preclinical study elucidated that there is a dynamic interplay between PI3K/AKT/mTOR pathway and AR signaling during the development of resistance to anti-androgens. [114]

1.5 Drug Discovery and Development

The process of drug discovery and development is generally time-consuming and expensive. A study conducted by DiMasi *et al* suggested that the average time and expenditure involved in discovering and developing a novel therapeutic agent may take up to 15 years and could cost approximately US $1 billion.[115] The standard drug discovery pipeline involves three critical stages 1) initial drug discovery which involves target identification and validation, hit identification, lead optimization 2) preclinical trials and 3) clinical trials (figure 1.10). [116]

The first step in the drug discovery process is identification and validation of a target. A drug target is a broad term which can be applied to a range of biological entities including proteins, genes and RNA. A good target needs to be efficacious, safe, meet clinical and commercial needs and most importantly it should be 'druggable'. In particular, A druggable target should be accessible to the putative therapeutic molecule and elicit a biological response which can be measured both *in vitro* and *in vivo*.[117] It is now known that certain target classes are more amenable to small-molecule drug discovery. One of the common target identification approaches is by examining mRNA/protein levels to determine whether they are expressed in

disease and if they are correlated with disease exacerbation or progression. An alternative approach is to use phenotypic screening to identify disease relevant targets.[118] Once identified, the target then needs to be fully prosecuted. Validation techniques range from *in vitro* tools through the use of whole animal models, to modulation of a desired target in disease patients.

Following the process of target validation, it is during the hit identification and lead discovery phase of the drug discovery process that compound screening protocols are utilized. The traditional and routine way of screening is through high-throughput screening, which involves the testing of the entire compound library directly against the drug target using a cell-based assay. Sometimes secondary assays may be needed to confirm the site of action of tested compounds.[119]

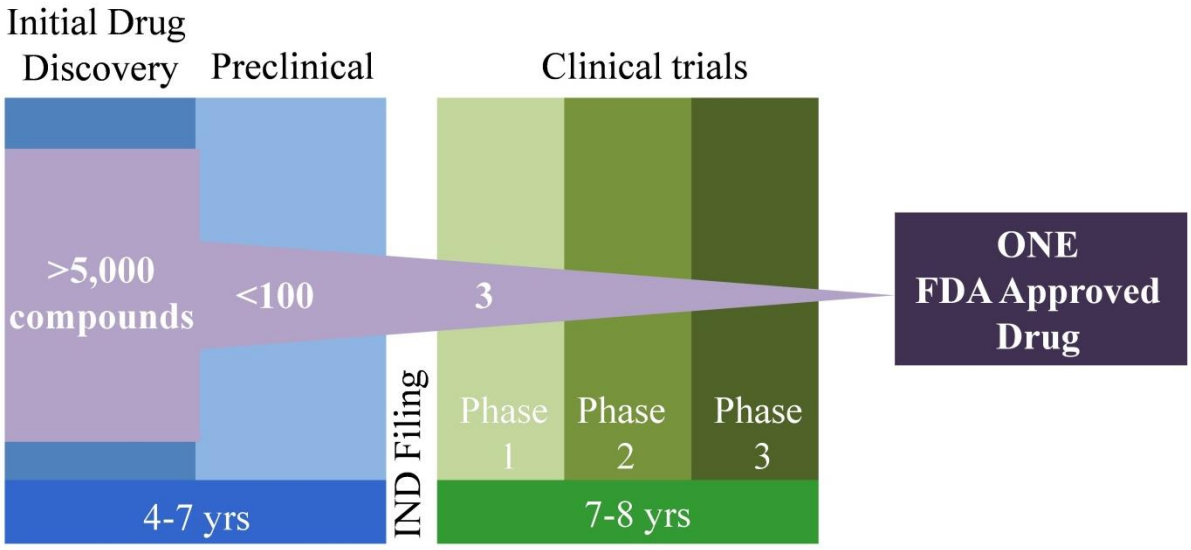


Figure 1.10. The typical drug discovery & development workflow and timelines involved.

Once the active compound is identified through the initial screening, a lead optimization stage takes place that involves fine-tuning undesired properties while maintaining its potency to develop better preclinical drug prototypes. The ultimate goal of preclinical studies is to accurately model the desired biological effects (efficacy and toxicity) of a drug-like candidate in animals in order to predict treatment outcome in patients. Currently, the outcomes of preclinical studies are poor with low success rate.[120, 121] The quality of the hit-to-lead starting point and the expertise of the available team are the key determinants of a successful outcome of "hit generation to preclinical" phase. The clinical trials (Phase1, 2 and 3) take up to 8 years and cost around 200 million USD.[122] It has been estimated that over 90% of drug candidates fail at this

stage. Due to this high attrition rate, new technologies should be incorporated into drug discovery process to enhance hit rate and lower the risk and costs.

Due to the higher estimates of time and costs involved in traditional approach, new avenues are needed to facilitate, expedite and streamline drug development process to save time and money. One promising strategy is to implement computer-aided drug discovery as a viable alternative and complement to high-throughput screening. It has been reported that large scale industrial high-throughput screening campaigns routinely used by pharmaceutical companies typically yield less than a 0.1% success rate. [123] In contrast, the use of a rationalized, computer-aided approach allows us to achieve a success rate as high as 50%.[124]

1.5.1 Computer-Aided Drug Discovery

The current post genomic era has been characterized by a large increase in a number of potential therapeutic targets amenable to investigation. In turn, this growth puts a pressure on pharmaceutical industry (and academic labs) to prioritize drug discovery programs and conduct them in a highly efficient manner. Therefore, computer-aided drug discovery (CADD) has emerged as an efficient approach to shorten the drug development cycle and save expenses. CADD is capable of increasing the success rate as it uses a much more targeted search than traditional high-throughput screening and combinatorial chemistry.[124] Moreover, it also explains the molecular basis of therapeutic activity. With the emergence of high computing facilities such as super computer clusters and graphics processing units, CADD has become a vital part of a modern drug discovery. Applications and benefits of CADD have been reviewed extensively and demonstrated in growing number of publications and supported by examples of drugs derived from the *in silico* approach.[125-127]

CADD is usually performed for the following three major purposes. (1) to quickly filter large databases into smaller sets of predicted active compounds that can be tested experimentally (2) to help in guiding optimization of lead compounds either to increase their target affinity or to optimize drug metabolism and pharmacokinetics properties or both (3) to facilitate the design of novel compounds either by growing template molecules one functional group at a time or by piecing together fragments into novel chemo-types.

CADD can be classified into two general categories. structure-based and ligand-based. The structure-based methods relies on the structural information of the therapeutic target (typically derived from X-ray crystallography, NMR or sometimes homology modeling) to

determine key protein-ligand interaction features. These methods include molecular docking and molecular dynamics simulations. The ligand-based approaches are generally preferred when no or little structural information is available and it mainly relies on the structure and bioactivity information of both known actives and inactive compounds. Commonly used ligand-based methods are similarity search, ligand pharmacophore and QSAR.

Molecular Docking. Docking recognizes and optimizes drug candidates by capturing molecular interactions between ligand and target macromolecules. The process involves two basic steps. Prediction of the ligand conformation as well as its position and orientation within the proposed binding site (usually referred to as *a pose*) and then ranking a pose via a scoring function. Ideally, sampling algorithms should be able to reproduce the experimental binding mode and the scoring function should also rank it highest among all generated conformations. Glide,[128] eHiTS[129] and GOLD[130] are some of the examples of commercially available docking programs.

A common limitation of the molecular docking approach is it may be ineffective in obtaining a meaningful correlation between its predicted scores and bioactivities of compounds. In the current work, we developed a novel approach to overcome some of such limitations by combining structure-based and ligand based approaches. This methodology is discussed in detail in chapter 5.

Molecular Dynamic (MD) Simulations. MD is a computationally expensive tool that provides more accurate information on protein-ligand interactions. Usually it is performed in the presence of an aqueous environment and under physiological conditions (certain temp, pressure and pH). MD simulation calculates the trajectory of a protein-ligand system by the application of Newtonian mechanics. In simple terms, the MD technology allows recreation of 'cartoon-like' trajectories of movements of a compound in a protein cavity during their mutual molecular recognition and binding. Such trajectories allow isolation of various 'snapshots' of a target protein's structure corresponding to its state before, during and after- ligand binding.

The binding free energy calculation is a critical step as post processing on the MD simulation which generally provides better correlation with experimental data than docking scores. Despite limitations in the force field of computations and high computational demands, MD simulations have become an important tool in drug discovery.[131]

Similarity Search. It is one of the traditional and most widely applied approaches in cheminformatics. This approach is developed based on the "Similarity Property Principle' which states that '*similar molecules should have similar biological properties*'. In this method, an active compound is used as a template/query to search against a large pool of chemicals and as a result compounds with better activity and/or chemicals deviating structurally from original query are obtained. Molecular fingerprint-based similarity is a commonly used method in similarity searching. The fingerprint is defined as bit string where each bit position accounts for the presence or absence of a given feature in a molecule.[132] If the feature is present in a molecule the bit is set to '1' and if the feature is not present, it is set to '0' forming a distinctive and unique fingerprint profile for each chemical structure. Popular substructure fingerprints include the Molecular ACCess System/MACCS structural keys[133] and Barnard Chemical Information Ltd. fingerprint.[134] The similarity between two molecules is identified by comparing bit strings of molecules and quantified as Tanimoto coefficient (Tc).[135]

$$Tc=C / (A+B-C)$$

Where, A and B are the number of features present in compounds A and B, respectively, and C is the number of features shared by A and B. Hence, Tc quantifies the fraction of features common to A and B to the total number of features of A or B, where the C term in the denominator corrects for double counting of the features.

Quantitative Structure Activity Relationship (QSAR). This approach has been applied for decades for constructing relationships between physicochemical (or structural) properties of compounds and their biological activities. The goal of QSAR is to derive a reliable and meaningful statistical model to predict the activity and/or binding affinity of new chemical entities. The construction of QSAR model typically comprises of two main steps. (1) description of molecular structures and (2) multivariate analysis for correlating molecular descriptors with observed activities (endpoints). The descriptors explain the properties of compounds including steric, topological, geometrical and hydrophobic properties.[136] Recently, Cherkasov *et al* provided an overview of the state-of-the-art QSAR methods and its application in the field of drug discovery and medicinal chemistry.[137]

Pharmacophore Modeling. Conventionally, a pharmacophore is defined as the specific 3D arrangement of functional groups within a molecular framework that are indispensable to attach to an active site of a protein or bind to a macromolecule. A pharmacophore feature

includes Hydrogen bond donor, acceptor, positively and negatively charged groups, hydrophobic regions and aromatic rings. A Pharmacophore map can be generated by superposition of active compounds to identify their common features. Based on the pharmacophore map either *de novo* design or chemical database searching can be carried out.[138]

1.6 Objective and Rationale of the Study

Once PCa progresses to castration-resistant stage, existing androgen deprivation therapies including the most potent and recently approved drug, Enzalutamide are ineffective due to the development of drug resistance. Numerous factors are causative for this phenomenon, including mutations in the ligand binding pocket of the AR resulting in structural changes in the receptor, allowing it to regain agonist conformation. These events weaken protein-drug interactions and allow anti-androgens to promote the recruitment of coactivators by the AR, enhancing its transcriptional activity. Therefore, there is an urgent need to develop new types of anti-AR therapeutics that exhibit entirely different modes of AR inhibition. For instance, rather than blocking androgen binding to the AR, such new drugs could target regulatory sites on the receptor and prevent co-activator recruitment directly.

The novel strategy was to explore the BF3 pocket as an alternative drug target that will help avoiding resistant escapes by the AR. BF3 is a protein-protein interaction site and is essential for AR transcriptional activity by recruiting co-regulator proteins such as FKBP52 and Bag-1L and engaging in crosstalk with the adjacent AF2 site. Thus, targeting BF3 pocket provided potential to not only inhibit wild-type AR, but also clinically relevant mutant forms of AR that confer resistance to clinically used Bicalutamide and Enzalutamide. The discovery of AR BF3 inhibitors involved the hit identification, lead optimization and upcoming preclinical evaluation. Although the developed lead compound needs further preclinical assessment, it has already demonstrated all desired profiles and may serve as a prospective therapeutic for advanced and castration resistant PCa.

1.7 Thesis Layout

Chapter 1 provided the background information on PCa, structure and function of AR, drug resistance mechanisms and modern drug discovery approaches. Chapter 2 presents materials and methods applied in the current work. Chapter 3, 4 and 5 summarize results of the research project

aimed to develop novel AR inhibitors which have been published in 2 research articles as indicated in the preface. Chapter 6 includes summary and future directions.

Chapter 2: MATERIALS AND METHODS

All methods applied in chapter 3, 4 and 5 are summarized in this chapter.

2.1 *In Silico* Modeling

2.1.1 Preparation of the Protein Structure for Docking

AR crystal structures 2YLO (2.50Å resolution)[139] and 4HLW (2.50Å resolution)[64] were used in the current study to perform molecular docking. These structures were prepared using the Protein Preparation Wizard within Maestro 9.3 suite (Schrödinger, LLC). All solvent molecules have been deleted and the bond order for the ligand and protein has been adjusted. The missing hydrogen atoms have been added, and side chains have then been energy-minimized using the OPLS-2005 force field, as implemented by Maestro. The active site was defined by a 12Å box centered on the crystallographic ligands, **YLO** and **17W** for 2YLO and 4HLW, respectively. No van der Waals scaling factors were applied; the default settings were used for all other adjustable parameters.

2.1.2 Ligand Preparation

All the compounds presented in this thesis were imported into a molecular database using Molecular Operating Environment (MOE) version 2009.[140] Hydrogen atoms were added after these structures were 'washed' (a procedure including salt disconnection, removal of minor components, de-protonation of strong acids and protonation of strong bases). The following energy minimization was performed with MMFF94x force field as it is implemented by the MOE and the optimized structures were exported into Maestro suite in the SD file format.

2.1.3 Virtual Screening

All the prepared compounds were docked into BF3 site of AR crystal structures using Glide SP program[128] implemented in Maestro suite.[141] This program approximates a complete systematic search of the conformational, orientational, and positional space of the docked compound. In this search, an initial rough positioning and scoring phase that dramatically narrows the search space is followed by torsionally flexible energy optimization on an OPLS-AA non-bonded potential grid for a few hundred surviving candidate poses. The very best candidates are further refined via a Monte Carlo sampling of pose conformation, which is a crucial step for

obtaining an accurate docked pose. The final poses are scored according to GlideScore 2.5. The starting point for Glide scoring is the empirically based ChemScore function of Eldridge *et al*. [142] GlideScore 2.5 modifies and extends the ChemScore function as follows.

$$\Delta G_{bind} = C_{lipo\text{-}lipo} \sum f(r_{lr}) + C_{hbond\text{-}neut\text{-}neut} \sum g(\Delta r)\, h(\Delta\alpha) +$$
$$C_{hbond\text{-}neut\text{-}charged} \sum g(\Delta r)\, h(\Delta\alpha) + C_{hbond\text{-}charged\text{-}charged} \sum g(\Delta r)\, h(\Delta\alpha) +$$
$$C_{max\text{-}metal\text{-}ion} \sum f(r_{lm}) + C_{rotb} H_{rotb} + C_{polar\text{-}phob} V_{polar\text{-}phob} +$$
$$C_{coul} E_{coul} + C_{vdW} E_{vdW} + \text{solvation terms}$$

The lipophilic-lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged.

At the next step, molecules which had a high computational binding affinities were re-docked into the same binding cavity using the electronic high- throughput screening (eHiTS) docking module.[129] The eHiTS algorithm takes a "divide and conquer" approach to docking, by breaking ligands into rigid fragments and connecting flexible chains. Each fragment is systematically and exhaustively docked everywhere in the receptor active site. Matching fragments are then reconnected to generate the docked pose. The generated poses are refined by a local energy minimization in the active site of the receptor, driven by the scoring function (as mentioned above). Compounds which received eHiTS docking scores above a cutoff value were selected for further *in silico* refinement.

2.1.4 Consensus Scoring and Voting

The determined docking poses of selected compounds were evaluated by 1) calculating rigorous docking scores, defined by the LigX module of the MOE package, which accounts for receptor/ligand flexibility; 2) predicting *p*K*i* of protein-ligand binding using MOE SVL script *scoring.svl* to improve accuracy of prediction of energies of hydrogen bonds and hydrophobic interactions; 3) by computing the root mean square deviation (RMSD) between docking poses generated by Glide and eHiTS programs to quantify their docking consistency.

Based on these sorted output values from the above four procedures, each molecule would then receive a binary 1.0 vote for every 'top10% appearance'. The final cumulative vote (with the maximum possible value of 5) was then used to rank the training set entries. Based on the cumulative vote, we have selected the most highly voted molecules and have subjected their docking poses to visual inspection. After this final selection step, promising compounds were purchased and tested experimentally. Most of the compounds were purchased from commercial vendors such as Enamine, Vitas-M and Life Chemicals with purity ≥95%. Some of the synthetic derivatives were provided by our collaborator Dr. Robert N Young (Professor, Dept. of Chemistry, Simon Fraser University, Burnaby, Canada)

2.1.5 Similarity Searching

Instant JChem,[143] a 2D similarity search tool from ChemAxon, was used to search through the ZINC database[144] (18 million compounds) for structural analogs. Compounds with higher Tanimoto coefficients values with respect to the query structure were selected for further studies.

2.1.6 Molecular Dynamics Simulations

In order to obtain the molecular picture of lead BF3 inhibitors **VPC-13163, VPC-13566** and GARRPR peptide binding in the AR BF3 site, we performed the explicit solvent molecular dynamics (MD) simulations starting from their docking poses in the crystal structure (4HLW) as predicted by Glide. All MD simulations were performed with the CUDA accelerated Amber 14 program. AR LBD force field parameters were obtained from the ff14SB force field and the ligands (**13163**,**13566** and DHT) parameters came from generalized amber force field with charges derived from a RESP fit using an HF/6-31G* electrostatic potential calculated using the Gaussian 09 program.

The initial AR LBD structure was taken from the 4HLW crystal structure. **13163** and **13566** were docked to the BF3 site. Five Cl− counter ions were added to neutralize the protein's total charge, and then the resulting complex structure was solvated using TIP3P water model in a cubic box, with box edges lying 10 Å from the outermost atoms of the complex. The numbers of water molecules used to solvate the complex were 11303.

MD simulations were carried out within AMBER 14 on WestGrid facilities from Compute Calculation Canada (https.//www.westgrid.ca). Firstly, it is a two-step minimization procedure. The first-step energy minimization was performed on solvent with the protein-ligand

complex restrained for a 2000-step steepest decent minimization and then 3000 steps of conjugate gradient minimization. The second step minimization was performed on the whole system without restraints. After the energy minimization, the system was heated up from 0 to 298 K over 40 ps with a harmonic restraint weight of 500 kcal/(mol·Å^2) on the whole system, followed by a 60 ps density equilibrium simulation by restraining the solute with a harmonic restraint weight of 10 kcal/(mol·Å^2). Then production MD simulation was conducted for 100 ns without any restraints under the NPT ensemble condition at a temperature of 298 K and pressure of 1atm.

During the simulation, the periodic boundary conditions and particle-mesh Ewald for long range electrostatics were employed. Short range interactions were cut off at 10 Å, and bonds involving hydrogen were held fixed using SHAKE. The time step was set to 2 fs, and the snapshot was taken every 500 steps to record the trajectory.

2.1.7 Binding Free Energy Calculation

The GB/SA component within the AMBER package was employed to compute the binding free energy.[145] This method was performed in parallel by running a python script "MMGBSA.py.MPI". Snapshots at 10 ps intervals in production phase were extracted for energy calculation within the whole simulation time. Average binding free energy on all snapshots was used for analysis. Based on the hypothesis, binding free energy was divided into the binding energy in vacuum and solvation free energy. The former was calculated by molecular mechanical algorithm with sander program and included internal energy, vdw energy and electrostatic energy; the latter consisted of polar and nonpolar solvation free energy. Generalized Born (GB) model was applied to calculate the contribution of polar solvation free energy, whereas solvent accessible surface area (SA) method was used to evaluate the nonpolar part of solvation free energy. The entropic penalty was omitted in terms of the same simulation condition and considerable computational cost. GB/SA method estimates the total solvation free energy of a molecule, ΔG_{solv}, by assuming that it can be decomposed into the "electrostatic" and "non-electrostatic" parts.

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel}$$

where ΔG_{nonel} is the free energy of solvating a molecule from which all charges have been removed (i.e. partial charges of every atom are set to zero), and ΔG_{el} is the free energy of first

removing all charges in the vacuum, and then adding them back in the presence of a continuum solvent environment. The GB model approximates ΔG_{el} by the following formula.

$$\Delta Gel = -\frac{1}{2}\sum_{ij}\frac{qiqj}{fGB(rij,Ri,Rj)}\left(1-\frac{\exp[-KfGB]}{\varepsilon}\right)$$

where $r_{i\,j}$ is the distance between atoms *i* and *j*, the R_i are the so-called effective Born radii, and *f*GB() is a certain smooth function of its arguments. The electrostatic screening effects of (monovalent) salt are incorporatedvia the Debye-Huckel screening parameter k.

2.1.8 Dataset Preparation for QSAR Models

A dataset of 106 BF3 inhibitors with anti-AR activity (Active or 1 if IC_{50}<10uM, Inactive or 0 if IC_{50}>10uM) in the range of low to high micro-molar range are collected.[146] These compounds were built and prepared using the MOE 2012 program as described above. All the prepared compounds were docked into BF3 site of AR crystal structure (4HLW, 2.3Å resoluion)[64] using Glide SP program.[128] No constraints were applied, and all other adjustable settings were kept as default. Top ranked conformation for each compound was considered to calculate atom-pair distance dependent descriptors. AR BF3 residues falling in the range of 5Å, 7Å and 10 Å (figure 2.1A) are considered to calculate those descriptors. 10Å was defined as upper cut off because interactions between the protein-ligand are weak beyond this threshold except Leonard-Jones potentials.

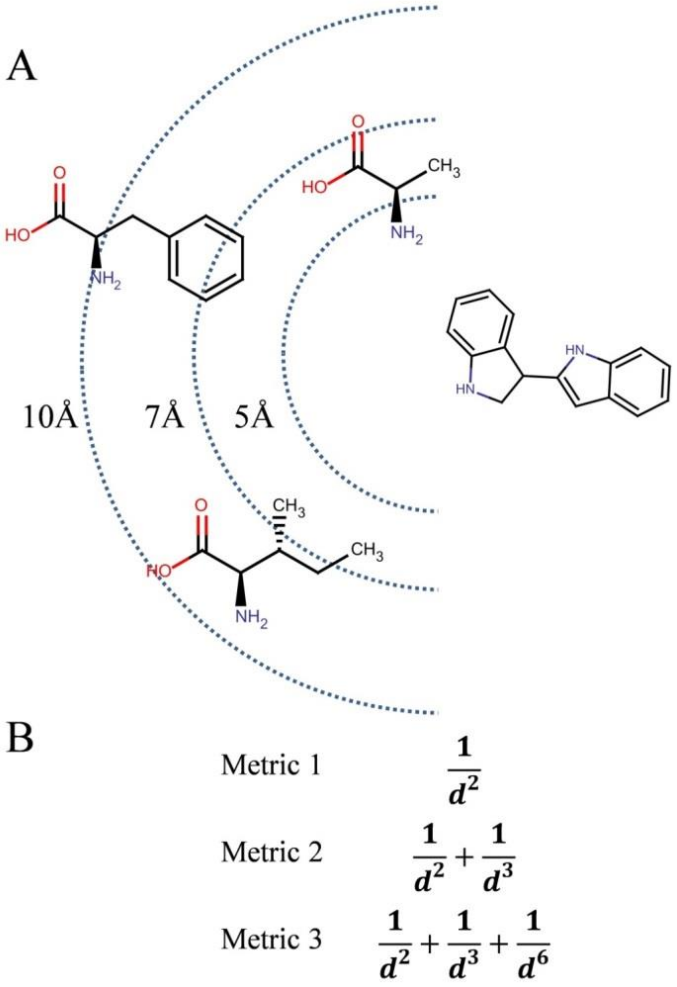


Figure 2.1. Methodology to develop protein-ligand atom pair descriptors. A) Distance-binding scheme for atom pair descriptors. B) Different type of metrics used to capture protein-ligand interactions.

2.1.9 Development of Descriptors

The sum of atomic distances between specific atomic type of BF3 residues and the docked ligands are captured as APDD (atom-pair distance dependent) descriptors. The atomic type was determined as 1) General.- based on geometry or hybridization and 2) Specific.- as per SMARTS atom typing scheme [147] (table 2.1). Specific atom type is incorporated because Kramer *et al* have recently shown that they yield significantly better statistical values compared to all other reported scoring functions.[148] Three different metrics (see figure 1B) are used to calculate the distances. This is implemented to see if change of precision plays a role in the final models.

General element typing. First, atoms were categorized according to their geometry and formal charge. For example, Nitrogen can be categorized into five types. Nsp^3, $N+sp^3$, Nsp^2, N+ sp2 and .Nsp2. Similarly, MOE distinguishes 3 atom types for carbon, 2 atom types for oxygen, and one atom type each for H, S, F, Cl, Br, and I. Overall, there are 16 different atom types for AR BF3 pocket and compounds. All the descriptors were calculated using SVL script encoded in MOE 2012.[140] Total number of calculated descriptors is 768 = 16 (ligand) X 16 (protein) X 3(number of bins).

Specific element typing. The atom typing scheme applied distinguishes 29 atom types for carbon, 17 atom types for nitrogen, 14 atom types for oxygen, 16 atom types for hydrogen, 3 atom types for sulfur, and one atom type each for F, P, Cl, Br, and I. Overall there are 84 atom types for the ligand atoms. Supporting Table 1 shows all carbon atom types, including the log*P* increments and the assignment of donor/acceptor/neutral as an example. Total number of calculated descriptors is 9828 = 84(ligand) * 39(protein) * 3(number of bins).

2D descriptors. In addition, about 300 descriptors such as the Molecular weight, molar refractivity and logP, were calculated as implemented in the MOE 2012.

Since the number of descriptors calculated is large, descriptor pruning is carried out. Descriptors with a variance below 0.01 and having a correlation above 0.90 are removed to reduce the signal-to-noise ratio.

Table 2.1. Sample of the atom typing SMARTS list for Carbon atoms used for the APDD descriptors.

#ID	SMARTS	Log P	Don/Acc/Neu	MR
C1	[CH4]	0.1441	N	2.503
C1	[CH3]C	0.1441	N	2.503
C1	[CH2](C)C	0.1441	N	2.503
C2	[CH](C)(C)C	0	N	2.433
C2	[C](C)(C)(C)C	0	N	2.433
C2	[CH3][N,O,S,F,Cl,Br,I]	-0.2035	N	2.753
C3	[CH2X4][N,O,S,F,Cl,Br,I]	-0.2035	N	2.753
C28	[C](=[O,N])[O,N]	-0.2783	N	5.007

#ID	SMARTS	Log P	Don/Acc/Neu	MR
C4	[CH1X4][N,O,S,F,Cl,Br,I]	-0.2051	N	2.731
C4	[CH0X4][N,O,S,F,Cl,Br,I]	-0.2051	N	2.731
C5	[C]=[!C;A;!#1]	-0.2783	N	5.007
C6	[C;A]=C	0.1551	N	3.513
C7	[CX2]#[A;!#1]	0.0017	N	3.888
C8	[CH3]c	0.08452	N	2.464
C9	[CH3]a	-0.1444	N	2.412
C10	[CH2X4]a	-0.0516	N	2.488
C11	[CHX4]a	0.1193	N	2.582
C12	[CH0X4]a	-0.0967	N	2.576
C13	[cH0]-[A;!C;!N;!O;!S;!F;!Cl;!Br;!I;!H]	-0.5443	N	4.041
C14	[c][#9]	0	N	3.257
C15	[c][#17]	0.245	N	3.564
C16	[c][#35]	0.198	N	3.18
C17	[c][#53]	0	N	3.104
C18	[cH]	0.1581	N	3.35
C19	[c](.a)(.a).a	0.2955	N	4.346
C20	[c](.a)(.a)-a	0.2713	N	3.904
C21	[c](.a)(.a)-C	0.136	N	3.509
C22	[c](.a)(.a)-N	0.4619	N	4.067
C23	[c](.a)(.a)-O	0.5437	N	3.853
C24	[c](.a)(.a)-S	0.1893	N	2.673
C25	[c](.a)(.a)=[C,N,O]	-0.8186	N	3.135
C26	[C](=C)(a)[A;!#1]	0.264	N	4.305
C26	[C](=C)(c)a	0.264	N	4.305
C26	[C](=C)a	0.264	N	4.305
C26	[C]=c	0.264	N	4.305
C27	[CX4][A;!C;!N;!O;!S;!F;!Cl;!Br;!I;!#1]	0.2148	N	2.693

#ID	SMARTS	Log P	Don/Acc/Neu	MR
CS	[#6]	0.08129	N	3.243

2.1.10 Mathematical Methods

To build QSAR models, we applied binary models implemented in WEKA[149] such as Decision Table, OneR, ADTree, BFTree, J48, BayesNet, SMO (SVM), NaiveBayes, Rotation Forest, Random Forest, Multilayer Perceptron, IB1, IBk, Kstar, Bagging, LogitBoost and Decorate. (http.//www.cs.waikato.ac.nz/ml/weka/). The classifier 'Relief Attribute Evaluation' with a search method 'Ranker' has been applied to rank descriptors. This evaluator assesses the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. Based on the ranking we selected top 10, 15 and 20 descriptors calculated by three metrics.

2.1.11 Evaluation of QSAR Models

Validation is the process by which reliability and relevance of a procedure are established for a specific purpose.[150] For validation of QSAR models, three strategies were adopted including 10-fold cross-validation [151], validation on external set and evaluation of the corresponding applicability domain (AD) criteria.

Once the 10 fold cross-validation was applied, all the models were ranked according to the Specificity, Sensitivity, Accuracy and Positive Predict Value (PPV). Sensitivity and specificity are common statistical measures of the performance for a binary classification research.[152] The sensitivity measures the proportion of active compounds which are correctly identified; and the specificity measures the proportion of inactive compounds which are correctly identified. Accuracy and PPV estimates the performance of the built QSAR classification models.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{FN} + \text{TN})}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Here, TP is the number of true positives; FP the number of false positives, TN the number of true negatives and FN is the number of false negatives predicted by the QSAR model.

In addition, root mean squared error (RMSE) is used to assess the quality of the relative predictions. RMSE measures the absolute accuracy of the prediction, *i.e.*, how well the outcome value is reproduced by the model.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{i,pred} - Y_{i,meas}\right)2}$$

The receiver operating characteristic (ROC) curve is employed to graphically present the model behavior in a visual way. A ROC curve shows the separation ability of a binary classifier by iteratively setting the possible classifier threshold.[152] As a result, a plot of the trade-off between the sensitivity (y-axis) and 1-specificity (x-axis) is shown. If the plot has a surface area of 1, a perfect classifier is found, and if the area equals 0.5, the classifier has no discriminative power at all.

2.1.12 External Set

Based on the chemical scaffold **13163**[146] an external library of derivatives was built. Approximately 400 derivatives were designed and prepared and docked into BF3 site as mentioned above. Based on the prediction of selected QSAR models and synthetic feasibility, twenty five compounds are selected for chemical synthesis.

2.2 Experimental Validation

All *in vitro* assays are performed by the team members of Dr. Paul Rennie (Professor, Dept. of Urologic sciences, University of British Columbia).

2.2.1 Cell Culture

LNCaP and PC3 human PCa cells were obtained from the American Type Culture Collection and grown in RPMI 1640 medium supplemented with 5% fetal bovine serum (FBS) (Invitrogen). LNCaP and PC3 cells were tested and authenticated by Idexx Radil (case no. 14616-2011) in June 2011. The LNCaP eGFP cells[153] were grown in phenol-red-free RPMI 1640 medium supplemented with 5% charcoal-stripped serum. HeLa-AR cells stably expressing the wild-type AR were grown in Dulbecco's modified Eagle's medium supplemented with 5% FBS. Enzalutamide-resistant LNCaP cells (MR49F) were provided by Dr. Zoubeidi[154] and were cultured in RPMI 1640 medium supplemented with5%FBS and 10 mM Enzalutamide. All cells were maintained at 37C in 5%CO2.

2.2.2 eGFP Cellular AR Transcription Assay

AR transcriptional activity was assayed as previously described by Tavassoli *et al.*[153] Briefly, stably transfected eGFP-expressing LNCaP human prostate cancer cells (LN-ARR2PBeGFP) containing an androgen-responsive probasin-derived promoter (ARR2PB) were grown in phenol-red-free RPMI 1640 supplemented with 5% CSS for 5 days. The cells were then seeded into a 96-well plate (35,000cells/well) and treated the next day with 0.1nM R1881 and increasing concentrations (0-100μM) of compound. After 3 days of treatment the fluorescence was measured (excitation, 485 nm; emission, 535 nm).

2.2.3 Prostate Surface Antigen assay

The evaluation of PSA excreted into the media was performed in parallel to the eGFP assay using the same plates (see above description). After the cells were incubated for 3 days 150μl of the media was taken from each well, and added to 150μl of PBS. PSA levels were then evaluated using Cobas e 411 analyzer instrument (Roche Diagnostics) according to the manufacturer's instructions.

2.2.4 MTS Assay

Cell proliferation was determined using the MTS cell proliferation assay following incubation with the compound (0-100μM) over 72h (CellTiter 961 Aqueous One Solution Reagent, Promega). In brief, 30μL of the reagent was added to cells in each well of the 96-well plate containing 200μL of media and incubated for 90 minutes at 37°C in 5% CO_2. The production of formazan was measured at 490 nm.

2.2.5 Bio-Layer Interferometry (BLI) Assay

The direct reversible interaction between small molecules and the AR was quantified by BLI using OctetRED (ForteBio). The LBB of the biotinylated androgen receptor (bAR) was produced in situ with AviTag technology.[155] The AviTag sequence (GLNDIFEAQKIEWHE) followed by a six residue glycine serine linker (GSGSGS) was incorporated at the N-terminus of the AR LBD (669-919). Escherichia coli BL21 containing both biotin ligase and AR LBD vectors were induced with 0.5mMisopropyl-β-D-thiogalactopyranoside (IPTG) in the presence of dihydrotestosterone (DHT) and biotin at 16°C overnight. The bacteria were then lysed by sonication, and the resulting lysate was purified by immobilized metal ion affinity chromatography (IMAC) with nickel_nitrilotriacetic acid (Ni_NTA) resin and cation-exchange chromatography (HiTrap SP).

Purified bAR LBD (50μg/mL) was bound to the super-streptavidin sensors over 50 min at room temperature. The sensor was kept in assay buffer [20 mMN-2-hydroxyethylpiperazine-N0-2-ethanesulfonic acid (HEPES), 150mM NaCl, 500μM tris(2-carboxyethyl)phosphine (TCEP), 500nM DHT, and 1% dimethylsulfoxide (DMSO)]. In all experiments, a known AF2-interacting peptide was used as a control to confirm functionality of the bAR LBD.

2.2.6 X-ray Crystallography of VPC-9002

All crystallographic experiments (pdb-4HLW) have been carried out as contract research by Structure-Based Design, Inc. (www.strbd.com).

2.2.7 *In vivo* Studies of Maximum Tolerated Dose (MTD) and Pharmacokinetics

In vivo experiments are performed by the team members of Dr. Emma Guns (Associate Professor, Dept. of Urologic Sciences, University of British Columbia). All animal experiments were conducted in accordance with the University of British Columbia Committee on Animal Care. For the Maximum Tolerable Dose (MTD) and serum level evaluation, twenty seven 6-8 week-old athymic nude mice (Harlan Sprague Dawley, Inc.) were intravenously, intraperitoneally or orally administered 50, 100 or 200 mg/kg of tested BF3 inhibitors solution formulated using 1.10, hydroxypropyl-cyclodextrin.ddH2O. Mice were monitored for 24 hours for signs of acute toxicity including death, lethargy, blindness and disorientation. In order to measure serum drug levels, tail bleed samples were taken from mice (100mg/Kg, n=3) following the administration at time points corresponding to 0, 1/2, 1, 2, 4, 6, 8 and 24 hours. Serum was separated by centrifugation (5min 3000G) and stored at -20 °C pending analysis by LC-MS.

Assessment of *in vivo* tumor growth for castration-resistant LNCaP xenografts and Enzalutamide-resistant (MR49F) xenograft transplantation. LNCaP. 6-8 week-old nude mice (Harlan Sprague Dawley, Inc.) weighing 25-31 g were subcutaneously inoculated with LNCaP cells (106 cells in BD matrigel, BD Biosciences, New Jersey, USA) at posterior dorsal site. Tumor volume, body weight, and serum PSA levels were measured weekly. When serum PSA levels reached more than 25 ng/mL, mice were castrated. When PSA recovered to pre-castration levels, mice were randomized into 3 treatment groups; vehicle, 10 mg/kg of Enzalutamide or 200 mg/kg of **VPC-13163**, 100 mg/kg of **VPC-13566** and treated orally twice daily for 3 weeks. Calipers were used to measure the three perpendicular axes of each tumor. The formula V= (L*W*H) π/6, where L is the length, W the width, and H the height, was used to calculate the tumor volume. Mice were also weighed weekly and monitored daily for signs of toxicity including death, lethargy, blindness and disorientation.

Enzalutamide-resistant (MR49F) cells. LNCaP-derived MR49F cells[154] were excised and transplanted to castrated mice treated with 10 mg/ kg of Enzalutamide daily. When tumors reached 100-150 mm3, the Enzalutamide treatment was stopped and mice were randomized into 3 treatment groups; vehicle, 10 mg/kg of Enzalutamide or required doses of BF3 inhibitors and treated orally twice daily for 3 weeks.

Chapter 3: DEVELOPMENT OF 2-((2-PHENOXYETHYL) THIO)-1H-BENZIMIDAZOLE DERIVATIVES

3.1 Background

As previously mentioned Fletterick and his group conducted a high throughput screen of AR with a known compound library and reported several compounds that bind to the receptor's surface at a novel structural pocket, referred to as BF3 site.[63] X-ray structural analysis revealed that chemicals such as flufenamic acid (FLUF), 3,3′,5-triiodo thyroacetic acid (TRIAC) and its derivative T3 (figure 3.1) were reported to bind to the BF3 cleft and interfere with AR activity. While these compounds revealed the importance of the BF3 site, they had weaker potency (IC_{50} > 50 μM) and were found to bind nonspecifically to the AR at multiple sites. Further investigational studies confirmed that BF3 is a protein-protein interaction site and is essential for AR transcriptional activity by recruiting co-regulator proteins such as FKBP52[65] and Bag-1L[66] and engaging in crosstalk with the adjacent AF2 site.

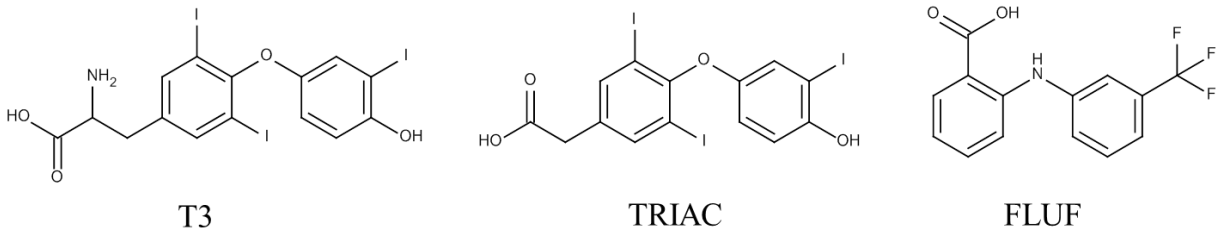


Figure 3.1. Chemical structures of small-molecule inhibitors reported to be able to bind to BF3 pocket as reported by Fletterick and co-workers. [63]

Based on the above observations, we proposed that rationally designed compounds selectively engaging BF3 will modulate coregulatory recruitment in physiological settings including PCa. In an effort to improve target affinity and BF3 specificity, our group has performed a large-scale computational screen followed by a series of experimental validation. We have discovered a number of novel small-molecules that inhibit AR activity in micro-molar range.[139] A comparison of the four reported co-crystal structures of the selected BF3 inhibitors (PDB code. 3ZQT, 2YLO, 2YLP, 2YLQ) revealed that residues of BF3 site may undergo significant conformational changes upon ligand binding. On the basis of the elucidated structure-activity-relationship of our virtual screening hits and binding poses in the co-crystals, **2-((2-**

phenoxyethyl)thio)-1H-benzo[d]imidazole (VPC-4035) was derived and proposed as a proper starting point for further lead optimization (see figure 3.2).

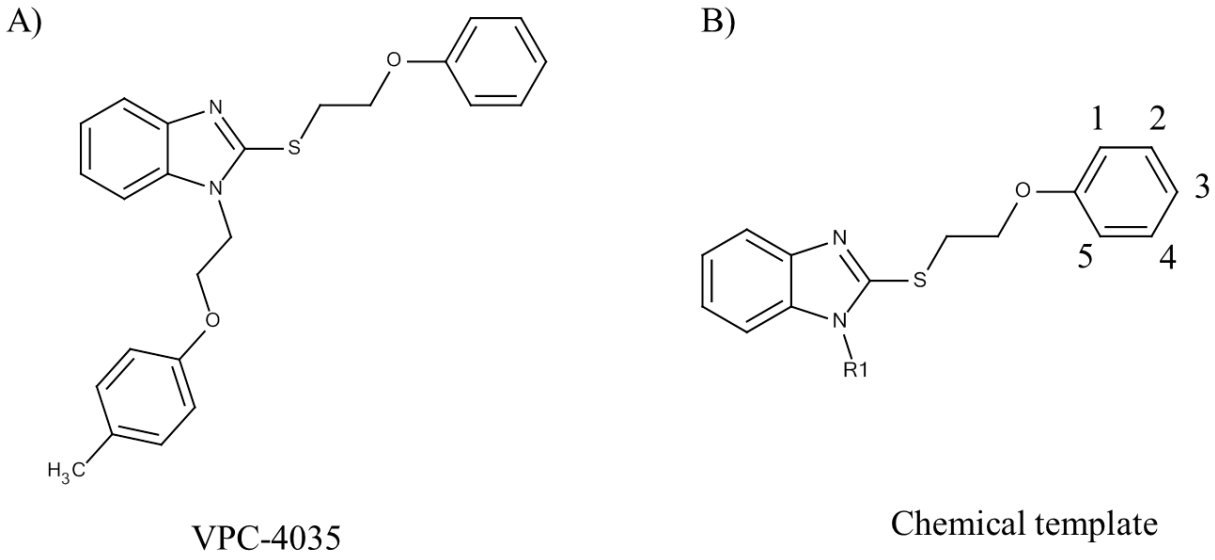


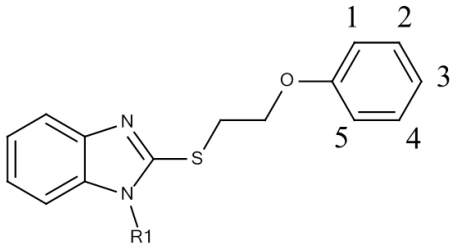
Figure 3.2. Known AR BF3 inhibitor and a chemical template derived from it. A) Previously identified AR BF3 inhibitor reported by Lack et al.[139] B) Chemical template used as a query to find analogues by 2D similarity search method.

3.2. Results

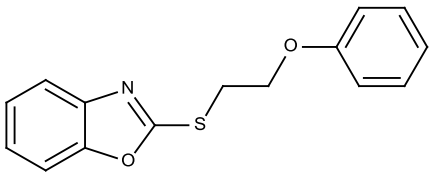
3.2.1 Identification of Analogues by 2D Similarity Search Method

Among previously identified BF3 inhibitors, (2-((2-phenoxyethyl) thio)-1-(2-(p-tolyloxy) ethyl)-1H-benzo[d]imidazole) derivative (compound **VPC-4035)** was selected as a lead candidate. A chemical template was designed based on the structure of **VPC-4035** (figure 3.2B) and a molecular similarity search was performed to identify compounds with different substitutions at R_1 and 1-5 positions of the benzene ring. Instant JChem,[143] a 2D similarity searching tool from ChemAxon, was employed to search through ZINC database 12.0.[144] All software parameters were set to their default values. A total of 30 ZINC compounds which generated Tanimoto coefficient above 0.6 with respect to the query structure were selected and tested for their anti-AR activity (see *analogues* section in Table 3.1).

Table 3.1. Structures and measured activities of the analogues of compound **VPC-4035** retrieved by 2D similarity search and proposed synthetic derivatives.



VPC-ID	AR Transcriptional IC_{50} (μM)	R_1	1	2	3	4	5
Analogues							
07	11.1	CH_2-C(O)O-iPr	H	H	CH_3	H	H
4068	12.7	C_2H_4OMe	H	H	CH_3	H	H
4075	13.7	CH_2-C(O)O-iPr	H	H	H	CH_3	H
4007	14.1	CH_2-C(O)OEt	H	CH_3	H	H	H
4034	22.8	C_2H_4O-Ph(4-Me)	H	H	H	OC_2H_5	H
4065	24.1	C_2H_4OMe	H	H	H	CH_3	H
4054	30.6	CH_2-C(O)O-iPr	H	H	CH_3	CH_3	H
4011	32.9	CH_2-C(O)OMe	H	H	CH_3	H	H
4101	35	Et	H	H	H	CH_3	H
4009	40.3	CH_2-C(O)OEt	H	H	CH_3	CH_3	H
4110	45	Me	H	H	H	CH_3	H
4069	55	$CH_2C(O)OEt$	H	H	H	C_2H_5	H
4012	163.8	CH_2-C(O)O-	H	H	CH_3	H	H
4003	>200	$CH_2C(O)$-N-Morph	H	H	H	CH_3	H
4006	>200	CH_2-C(O)OMe	H	H	H	C_2H_5	H
4008	>200	CH_2-C(O)OMe	H	H	CH_3	CH_3	H

VPC-ID	AR Transcriptional IC_{50} (μM)	R_1	1	2	3	4	5
4010	>200	CH_2-C(O)OEt	H	CH_3	H	CH_3	H
4020	>200	CH_2-C(O)O-	H	H	H	C_2H_5	H
4025	>200	C_2H_4-C(O)O-	H	H	CH_3	H	H
4028	>200	C_2H_4O-Ph	H	H	H	CH_3	H
4032	>200	CH_2-C(O)O-	H	H	CH_3	CH_3	H
4037	>200	CH_2-C(O)O-	H	H	H	CH_3	H
4039	>200	C_2H_4-C(O)O-	H	CH_3	H	H	H
4041	>200	C_2H_4-C(O)O-	H	H	H	CH_3	H
4052	>200	$CH_2C(O)NEt_2$	H	H	H	CH_3	H
4060	>200	CH_2-C(O)O-	H	CH_3	H	CH_3	H
4061	>200	CH_2-C(O)O-	H	CH_3	H	H	CH_3
4064	>200	CH_2-C(O)O-	H	H	H	t-Bu	H
4076	>200	CH_2-C(O)O-Me	H	$CH_2CH=CH_2$	H	H	H
4096	>200	C_2H_4-C(O)O-	H	H	H	C_2H_5	H
Synthetic Derivatives							
9002	4.2	H	H	H	H	H	H
9045	12	Et	H	H	H	H	H
9047	>200	C_2H_4OH	H	H	H	H	H
9050	>200	C_2H_4COOH	H	H	H	H	H
9055	>200	$CH_2CH=CH_2$	H	H	H	H	H
9100	>200						

3.2.2. Cell-Based Testing and *In Vitro* Characterization

The selected compounds were screened for their ability to inhibit AR transcriptional activity using a nondestructive, cell-based enhanced green fluorescent protein (eGFP) AR transcriptional assay [153] (see chapter 2. materials and methods). In this assay, the expression of eGFP is under the direct control of an androgen responsive probasin-derived promoter and enables quantification of AR transcriptional activity. 13 of the purchased compounds exhibited >50% inhibition of AR transcription at a concentration of 50μM. These were subjected to concentration-dependent titration to establish corresponding IC_{50} values (table 3.1). The observed IC_{50} values were estimated to be in the range of 11-60μM. These inhibitors were then tested in SRC2 peptide and androgen displacement assays (using commercial kits available from Life Technologies) for their ability to displace SRC2 peptide from the AF2 site and androgen from the ABS, respectively. None of the compounds were active in these assays, confirming that they target the BF3 pocket. From the cell proliferation assay, the compounds were also confirmed to be nontoxic to non-AR containing cells at a 50μM concentration administered for over 72h.

Since the molecules selected from the 2D similarity search did not result in compounds with improved cell-based activity, the lead optimization was initiated.

3.2.3 Rational Design, Synthesis and Characterization of 2-((2-phenoxyethyl) thio)-1H-benzimidazole (9002)

Based on the crystallographically determined binding pose of a compound **4035** (2YLO) and the corresponding activity profiles of their analogues, we hypothesized that solvent-exposed substituents at the R_1 position of the ligands are not likely to contribute to target affinity. To test this hypothesis, we designed a compound **9002** where R_1 = H (table 3.1). The structure was built using the MOE program and energy minimization was performed by applying the MMFF94X force field. The compound was then docked into the AR crystal structure (2YLO structure) using the Glide SP program without applying any constraints. From the docking pose, we could observe that compound **9002** is anchored to the protein site by a hydrogen bond that it forms with the Glu837 side chain. Compound **9002** was synthesized and evaluated by the eGFP transcriptional assay. As anticipated, it exhibited approximately a 3-fold increase (IC_{50}=4.2μM) of AR transcriptional activity inhibition compared to the parent substance (IC_{50}=13.1μM). The dose-response curve for compound **9002** is presented on figure 3.3A.

Furthermore, we have tested compound **9002** with the AF2 peptide displacement assay and androgen displacement assay where it did not demonstrate any detectible levels if activity, confirming that it is a specific BF3 binder. Biolayer interferometry (BLI) studies demonstrated a direct reversible interaction between this compound and AR LBD (figure 3.3B).

Based on these observations, we concluded that the formation of H-bond between NH of the benzoimidazole moiety of compound **9002** and the side chain of the Glu837 residue is a significant factor for protein-ligand affinity. This observation becomes particularly obvious when compound **9002** is compared with the synthetic analogues **9045**, **9047**, **9050** and **9055,** which were designed and tested as negative probes and which also received lower docking scores due to the loss of a critical H-bond with Glu837. Upon testing for anti-AR activity, all these derivatives except compound **9045** turned out to be completely inactive whereas compound **9045** showed a 3-fold decrease in activity (IC_{50}=12μM) in the eGFP assay (compounds shown as synthetic derivatives in table 3.1). Similarly, replacing benzoimidazole ring with benzoxazole, as in the case of compound **9100,** caused a drastic effect in activity and binding to the AR LBD. Since compound **9002** has a promising experimental activity profile, it was subjected to structural elucidation using x-ray crystallography as contract research by Structure-Based Design, Inc. (www.strbd.com).

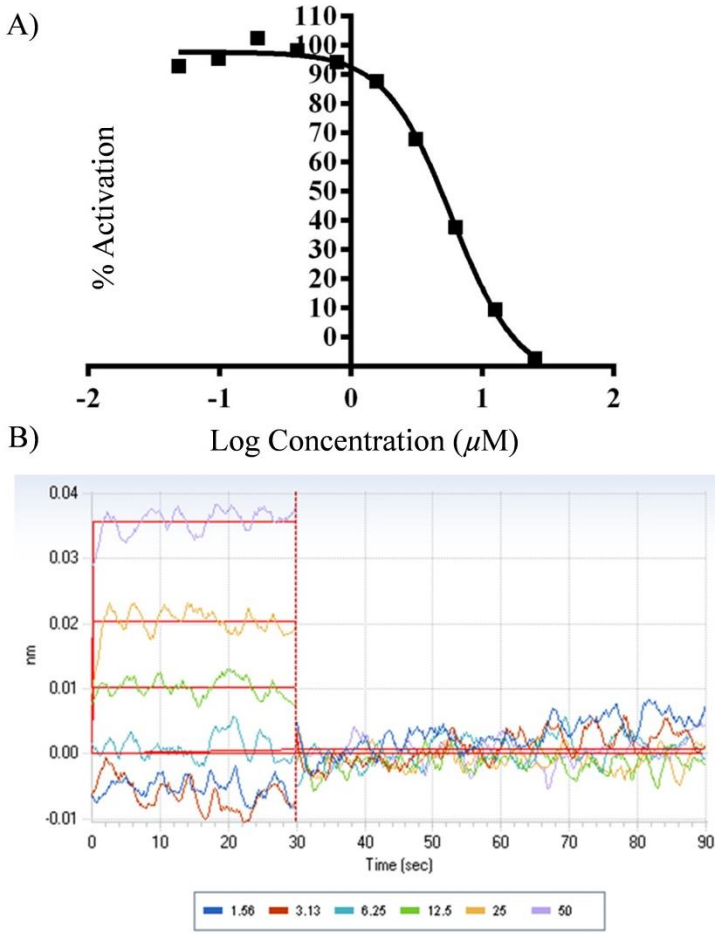


Figure 3.3. Activity profile of **9002**. A) Dose-response curve (0-25μM) illustrating the inhibiting effect of the compound **9002** on the AR transcriptional activity in cells. Data was fitted using log of concentration of the inhibitors vs % activation with GraphPad Prism 6. B) BLI dose-response curves (0-50μM) reflecting the direct binding of the compound **9002**.

3.2.4. Crystallographic Structure of AR in Complex VPC-9002

In an effort to unambiguously confirm the site of the compound's interaction, x-ray crystallographic studies were conducted with the AR and compound **9002**. Following optimization of the soaking protocol, the structure of the AR in complex with compound **9002** was determined to 2.5Å resolution. The crystallographic data refinement statistics for the corresponding PDB entry **4HLW** are presented in table 3.2. In the present crystallographic data set, there was clear electron density observed at the BF3 site, supporting the presence of the inhibitor. Interestingly, unlike the cases of previously published BF3 binders such as **4035**, [139]TRIAC, T3 and FLUF[63], compound **9002** was found to reside specifically in the BF3 site.

Since the crystallographic information is in agreement with the activity data, compound **9002** could be characterized as a specific BF3 inhibitor. Figure 3.4A shows a good structural fit of compound **9002** inside the target cavity. Notably, the experimentally determined configuration of the BF3-bound molecule is similar to its docking predicted pose generated by Glide (r.m.s.d=0.62Å), which gave confidence to rely on the adopted *in silico* protocol (supplemental figure 3.1).

The structure of the AR-LBD in complex with compound **9002** is generally similar to the previously reported structures 2YLO. Compound **4035** differs slightly in terms of its positioning inside the BF3. Thus, as predicted compound **9002** forms a strong hydrogen bond between the NH benzimidazole moiety and side-chain carboxyl of the Glu837. Moreover, this compound maintains strong hydrophobic contacts with the neighboring residues, including Ile672, Phe673 and Leu830. Additional stabilization of the protein-ligand complex occurs due to T-shaped arene-arene conjugation between the phenyl ring of compound **9002** and the Phe826 side chain. These interactions possibly explain the increased potency of compound **9002** (IC_{50} =4.2μM) compared to its parental compound **4035** as well as other structural analogues listed in table 3.1 (where the corresponding IC_{50} values range from 11μM to >200μM). Since compound **9002** demonstrated improved AR inhibitory activity profile and could be experimentally resolved inside the BF3 site, it was advanced into further optimization studies.

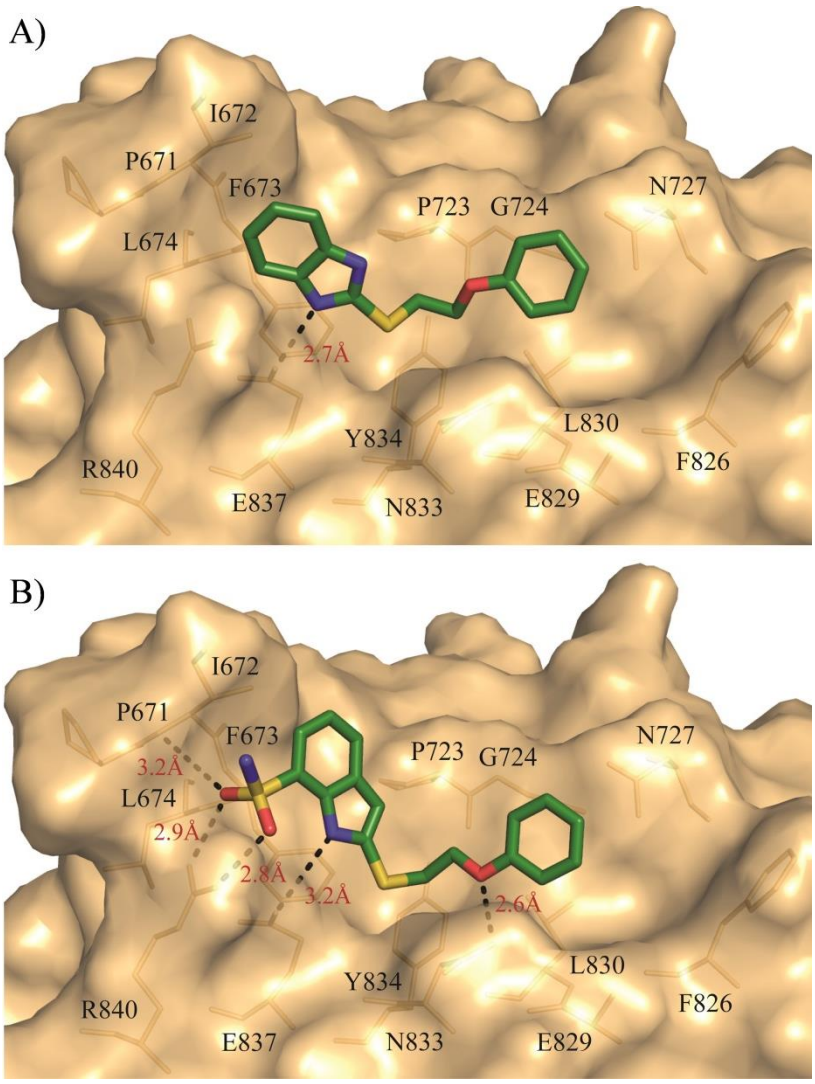


Figure 3.4. Binding orientation of benzimidazole derivatives. A) X-ray crystal structure of compound **9002** bound to BF3 pocket on the surface of human AR. Hydrogen bonds are shown in red. B) Binding orientation of compound **9145** inside the BF3. Hydrogen bonds are shown in black.

Table 3.2. Data collection and refinement statistics of 4HLW.

PDB Code	4HLW
X-ray Source	Synchrotron
Space Group	P212121
a, *b*, *c* (Å)	55.19, 66.30 and 73.01
α,β,γ (°)	90.0, 90.0 and 90.0
Data collection statistics	
Resolution (Å)	2.5
R_{sym} or R_{merge}	0.136/(0.580)
No. of unique reflections	12422/(1786)
I/σ (*I*)	7.01/(2.37)
Completeness (%)	99.93/(100)
Multiplicity	6.0/(6.2)
Refinement and model statistics	
Resolution (Å)	2.5
No. reflections used (work + test)	9717
R_{work}^{a}	0.188
R_{free}^{a}	0.250
No. of residues	244
No. of water molecules	13
Additional molecules	4
Total No. of atoms	1953
R.M.S.D bond length (Å)	0.029
R.M.S.D bond angles (Å)	1.21
Wilson B-factor (Å^2)	35.4
Mean B-factor (Å^2)	47
Ramachandran statistics (%)	

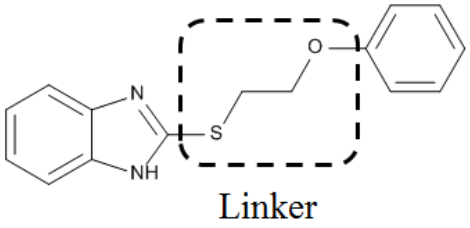
PDB Code	4HLW
Favored region	98.0
Additional allowed region	7.0
Generously allowed region	0.9
Disallowed	0

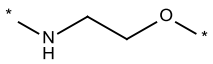
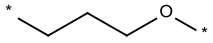
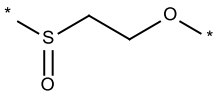
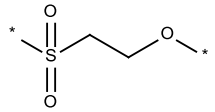
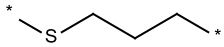
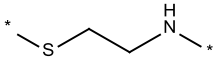
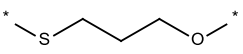
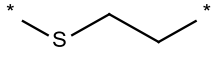
[a] R_{work} and $R_{free} = \Sigma_h||Fo(h) - F_c(h)||/\Sigma_h||$ for the working set and test set (5%) of reflections, where $Fo(h)$ and $Fc(h)$ are the observed and calculated structure factor amplitudes for reflection.

3.2.5. Structure Activity Relationship (SAR) for 2-((2-phenoxyethyl) thio)-1H-Benzimidazole as AR BF3 Ligands

To further explore the relevance of various structural elements of compound **9002**, we designed 8 of its close derivatives by substituting its sulfur and oxygen atoms in the linker region (table 3.3). Data obtained from Bio-Layer Interferometry (BLI) experiments and eGFP cellular assays revealed that subtle changes in the linker can have profound effects on target binding and inhibitory activity. Accordingly, we investigated the significance of a sulfur atom in compound **9002** by replacing it with nitrogen (compound **9105**) and carbon (compound **9130**). These modifications abolished the cellular activity of derivatives **9105** and **9130** and their binding to the AR. Similarly, the replacement of sulfur with sulfinyl and sulfonyl fragments caused a significant drop in activity of the corresponding derivatives **9058** and **9057**. According to the modeling data the loss in activity of **9058** and **9057** was due to the presence of oxygen which disrupts the critical van der Waals contacts with Phe673, Tyr834 of the BF3 pocket. Thus, our further investigation was focused around the oxygen atom of linker region in compound **9002**. In particular, the replacement of oxygen with carbon resulted in compound **9097**, which demonstrated a 4-fold drop in anti-AR potential (12μM). Surprisingly, replacing oxygen with nitrogen (compound **9131**) exhibited detrimental effects on its binding and activity. Increasing the length of the linker fragment from SC_2H_4O to SC_3H_6O resulted in loss of activity of a derivative **9061**, likely caused by its poor fit inside BF3. Shortening the linker region (*i.e.* removing oxygen atom and exchanging the SC_2H_4OPh fragment to SC_3H_6Ph) did not result in any major alteration of anti-AR potential, with the corresponding IC_{50} assessed for compound **9099** at 7.4μM.

Table 3.3. Structures and measured activities of the synthetic derivatives of compound **9002** with different linkers attached.

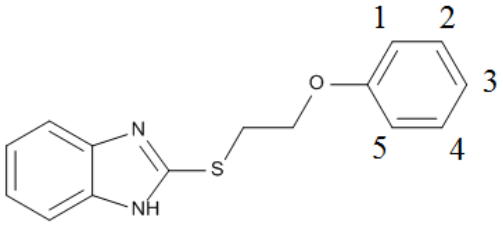


VPC-ID	AR Transcriptional IC_{50} (μM)	Linker
9105	>200	
9130	>200	
9058	>200	
9057	>200	
9097	12	
9131	>200	
9061	>200	
9099	7.4	

Another focal point of the study was evaluation of the effect of substitutions at the benzene ring of compound **9002** (table 3.4). The starting three analogues, **9112**, **9117** and **9006,** were designed and synthesized by introducing a methyl group at R_1, R_2 and R_3 positions of the core. As predicted, compound **9117** showed a 2-fold increase in AR-suppressing activity (IC_{50}=1.8μM) compared to the parental compound **9002** whereas molecule **9112** demonstrated 5-fold decrease while compound **9006** demonstrated only slightly lowered activity (IC_{50}=7.0μM). As our docking models demonstrated, the presence of a methyl group at the meta- position ensures additional hydrophobic contacts with the Phe826 and Leu830 residue of BF3 and contributing towards enhanced ligand binding (supplemental figure 3.2). Similarly, 2, 5-methyl substitution was well-tolerated and led to further enhanced activity of the corresponding derivative **9114** (IC_{50}=2.7μM). The dose response curve for compounds **9117** and **9114** is presented in figure 3.4A. The BLI experiment confirmed a direct reversible interaction between compounds **9117**, **9114** and the AR as shown in figure 3.5B and C. We have also explored the introduction of chlorine atom into R_2 (compound **9103**) and R_3 (compound **9101**) substitution positions, which appear not to affect the overall activity. A larger sulfonamide group was not tolerated at position R_3 (**9102**) and caused a 15-fold decrease in the inhibition activity.

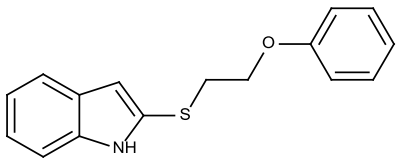
An effort was also made to replace imidazole with an indole moiety in the developed BF3 inhibitors (table 3.5). Initially, we have made an indole-based compound **9088** which demonstrated a similar level of AR inhibition compared to compound **9002** with a corresponding IC_{50} of 5.4μM. Derivative **9145** was designed by adding a sulfonamide group to the 7-position of compound **9088**. According to the docking model shown in figure 3.4B, sulfonamide forms additional networks of hydrogen bonds with nearby residues *i.e.* side chain of Arg840, Glu837, Asn833 and backbone of Pro671. As a result, the compound demonstrated 4-fold increase its in activity (figures 3.4A). The direct reversible interaction between this compound and the AR LBD was also detected by the BLI (Figure 3.5D). Based on the above observations, compounds **9002**, **9117**, **9114** and **9145** were selected for further evaluation.

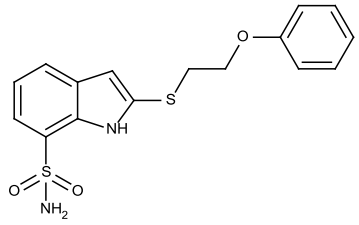
Table 3.4. Structure and activity data for synthetic derivatives with different substitutions around phenyl ring.



ID	AR Transcriptional IC_{50} (μM)	1	2	3	4	5
9112	19	CH_3	H	H	H	H
9117	1.8	H	CH_3	H	H	H
9006	7	H	H	CH_3	H	H
9114	2.7	CH_3	H	H	H	CH_3
9103	5	H	Cl	H	H	H
9101	4	H	H	Cl	H	H
9102	62	H	H	SO_2NH_2	H	H

Table 3.5. Structure and Activity Data for 2-((2-phenoxyethyl) thio)-1H-indoles.

ID	AR Transcriptional IC_{50} (μM)	Structure
9088	5.4	

ID	AR Transcriptional IC_{50} (μM)	Structure
9145	1.5	

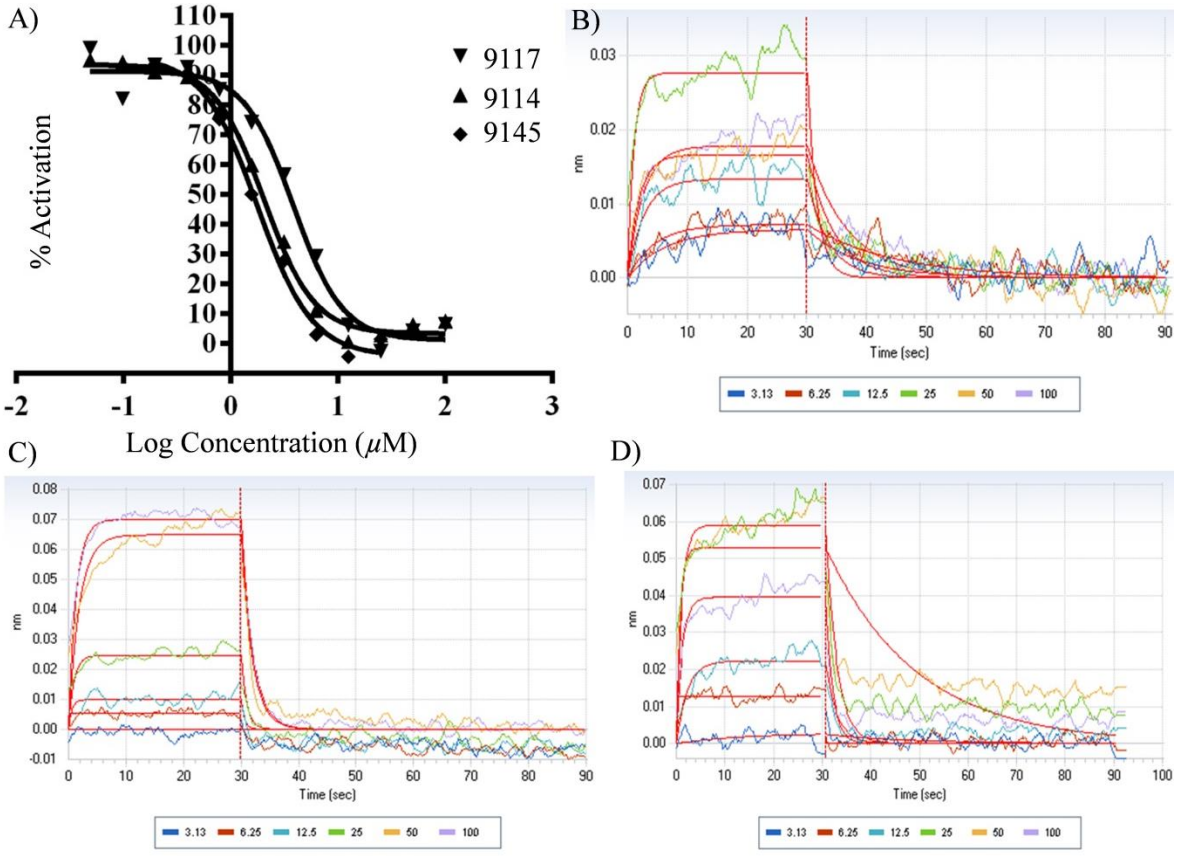


Figure 3.5. Activity profile of benzimidazole derivatives. A) Dose-response curves (0-100µM) illustrating the inhibiting effect of the compounds **9117, 9114**, and **9145** on the AR transcriptional activity in cells. Data points represent the mean of two independent experiments performed in triplicate. Data was fitted using log of concentration of the inhibitors vs % activation with GraphPad Prism 6. BLI dose-response curves (0-50µM) reflecting the direct

binding of the compound B) compound **9117** C) compound **9114** and D) compound **9145** to the AR LBD protein.

3.2.6. Derivatives 9002, 9117, 9114 and 9145 Reduce PSA Expression in LNCaP and Enzalutamide- Resistant Cells

To rule out possible false positive hits in the AR transcriptional eGFP assay, we validated the activity of compounds **9002**, **9117**, **9114** and **9145** by quantifying their effect on the production of the prostate specific antigen (PSA) in prostate cancers cell lines. PSA is a serine protease whose expression is dependent on AR activity level in the cell.[156] PSA is widely used as a marker for PCa as its serum concentration is associated with this pathological condition. As expected, these derivatives induced a dose-dependent decrease in PSA levels in LNCaP prostate cancer cells [157] with corresponding IC_{50s} value determined as 4.3, 3.3, 1.9 and 1.6μM respectively (Figure 3.6A). These compounds were also evaluated using *in house* developed Enzalutamide- resistant PCa cell line (MR49F cells).[154] These AR inhibitors are significantly more effective than Casodex and Enzalutamide in these cells. Figure 3.6B demonstrates that anti-AR drugs are ineffective, with IC_{50}s greater than 100μM. On the other hand, even though compound **9002** (IC_{50}=21μM) reduced PSA levels moderately, derivatives **9117**, **9114** and **9145** (IC_{50}=13, 6.8 and 6.4μM) were quite effective in these MR49F cells. Hence, the inhibition values obtained for inhibition of PSA in LNCaP and MR49F cells confirms the effectiveness of these inhibitors on the AR signaling pathway.

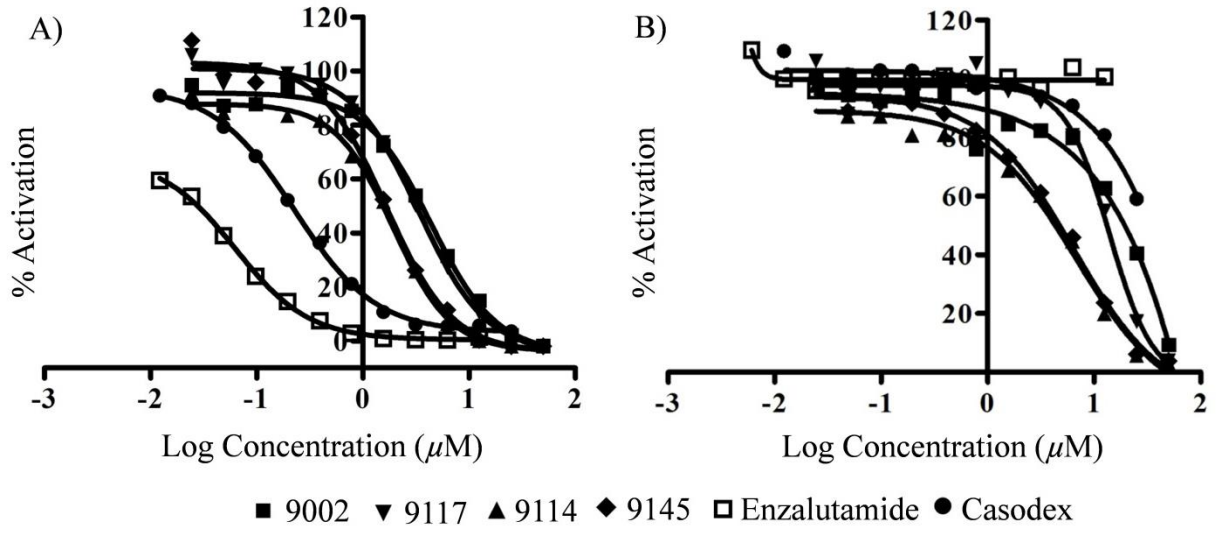


Figure 3.6. Inhibition effect of **9002, 9117, 9114** and **9145** in comparison to Casodex and Enzalutamide on PSA in dose response manner in A) LNCaP cells B) MR49F cells.

3.2.7. Derivatives 9002, 9117, 9114 and 9145 Reduces Cell Growth in LNCaP and Enzalutamide- Resistant Cells

To ascertain the growth inhibitory potential of AR inhibitors **9002**, **9117**, **9114** and **9145** we evaluated their ability to inhibit growth of LNCaP[157] and MR49F cells[154] as well as on AR-independent PC3 cells.[158] The cell viability was assessed after 4 days of incubation with the test compounds at a concentration of 6μM. Figure 3.7 shows that compound **9002** did not have any significant inhibition effect on these cancer cells whereas its derivatives **9117, 9114** and **9145** suppress cancer cells quite effectively at the concentration measured. Derivatives **9114** and **9145** exhibit a particularly strong effect on the growth of both LNCaP and MR49F cells. Moreover, derivatives **9117, 9114** and **9145** did not show any effect on AR independent PC3 cell lines, confirming their AR-specific activity.

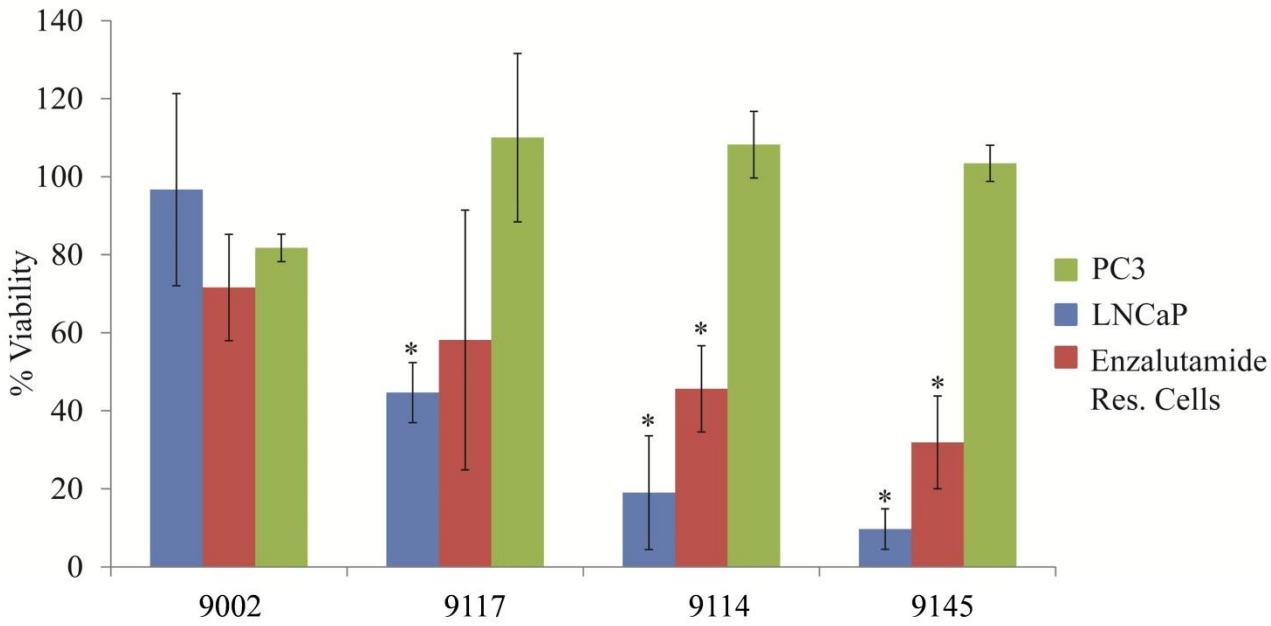


Figure 3.7. The effect of compounds **9002, 9117, 9114** and **9145** on cell viability in LNCaP, Enzalutamide resistant cells and PC3 cells. % cell viability is plotted at 6µM concentration. Data are presented as Mean ± SEM. A *p* value <0.05 was considered very significant effect on LNCaP and Enzalutamide resistant cells compared with PC3 cells.

3.2.8. Derivatives 9002, 9117, 9114 and 9145 are Selective AR BF3 inhibitors

We undertook to profile the selectivity of these derivatives for AR over ERα, other member the steroidal nuclear receptor subfamily. The compounds were tested for their ability to inhibit 17βEstradiol-ERα-mediated gene transcription in MCF-7 human breast cancer cells using luciferase reporter whose expression is driven by consensus estrogen response element. Supplementary figure 3.3 shows that the compounds do not inhibit ERα transcriptional activity compared to Tamoxifen measured at 3 different concentrations (10, 5 and 1µM). This confirms that these inhibitors are AR BF3 specific.

3.3 Discussion

Surface pockets or protein-protein interaction sites are often considered as attractive options for therapeutic targeting. However, identifying small-molecules that modulate these sites is often difficult owing to issues such as lack of a well-defined deep binding pocket. Although surface sites are challenging drug targets, their adaptive character can provide binding grooves for compounds and thus opportunities for drug discovery.[159, 160] In the case of the AR, targeting its

BF3 pocket offers a promising alternative strategy to create novel therapeutics for castration-resistant PCa. Since the AR BF3 is surface exposed, identifying compounds with significant activity profiles and developing structure-activity relationship around them is challenging.

We have previously utilized the power of virtual screening combined with experimental evaluations to discover a number of small molecules that effectively target the BF3 site of the AR. On the basis of one of the identified inhibitors (**4035**), we developed a series of analogues with improved anti-AR activity. In particular, a simplified yet more active derivative **9002** was synthesized, experimentally evaluated and crystallographically resolved inside the AR BF3 target cavity. The reported structure 4HLW demonstrated that bezinoimidazole moiety of the parental compound **9002** makes a strong H-bond with neighboring residue Glu837. The information obtained by both inhibition experiments and x-ray crystallography studies indicated that compound **9002** is a strong BF3-specific inhibitor. Hence, we synthesized a number of derivatives of this compound and explored their structure-activity relationship in the context of anti-AR potency**.**

We initially modified the linker region of compound **9002**. Replacement of the oxygen atom in SC_2H_4O was tolerated, but did not result in further improvement of potency. Other modifications completely abolished anti-AR activity and binding. Hence, we focused on introducing groups at the benzene ring of the compound **9002** template. By comparison, addition of small hydrophobic substituents such as methyl at various position of the benzene ring was able to enhance anti-AR potency of the corresponding derivatives of compound **9002**. In particular, compounds **9117** and **9114**, containing methyl at meta- and di-ortho positions, demonstrated IC_{50} in single digit μ-molar range. Replacement of the benzoimidazole moiety in compound **9002** with a synthetically more favorable indole fragment did not significantly alter the activity of the derivatives. The introduction of a sulfonamide group at 7-position of the indole core (compound **9145**) further increased the target affinity by providing additional hydrogen bonds with Arg840 and Phe673 residues.

These findings culminated in the discovery of rather potent AR inhibitors **9002, 9117, 9114** and **9145** with the corresponding IC_{50}s of 4.2, 1.8, 2.7 and 1.5μM respectively, which are 5-10 times lower than the IC_{50} of 13.1μM for the parental compound **4035**. The activity of these chemicals was further confirmed by their ability to decrease the levels of PSA in LNCaP and

Enzalutamide-resistant PCa cells. Compounds **9002, 9117, 9114** and **9145** exhibited IC_{50}s of 4.3, 1.9, 3.3 and 1.6μM respectively in LNCaP cells. Similar potencies were also observed in Enzalutamide-resistant cell line. Compounds **9114** (IC_{50}=6.8μM) and **9145** (IC_{50}=6.4μM) turned out to be especially effective in comparison with clinically used Casodex and Enzalutamide (IC_{50}>100μM in these resistant cells). The PSA inhibition figures were also in agreement with the above numbers giving further confidence in these BF3 inhibitor prototypes.

In summary, while we obtained 30 analogues of compound **4035** by 2D similarity search, but they were not very active. However, we further rationally developed, synthesized and tested 21 benzimidazole derivatives of compounds **4035** and 9 of them showed equivalent or improved potency against the AR. Similarly, we created and evaluated two indole derivatives which also exhibited enhanced anti-AR potency. These initial results obtained with indole-based compounds are encouraging and will be further investigated. Moreover, the structure of the AR in complex with compound **9002** (one of the synthetic derivatives) was elucidated and turned out to be in very good agreement with our prior predictions, providing additional confidence in our modeling approach.

Drug resistance remains a fundamental cause of therapeutic failure in cancer therapy.[161, 162] In PCa, cancer progression to a drug-resistant phenotype in the presence of an antagonist possibly through selection of cells with epigenetics or mutational changes that bypass the inhibitory action of the drug. Our lead derivatives were tested for their ability to inhibit AR in LNCaP PCa cell lines including those which have developed resistance to the recently approved potent anti-androgen, Enzalutamide.[71] Results from cell viability assays indicated that the tested derivatives exhibited effective inhibition of growth in both LNCaP and Enzalutamide-resistant cell lines containing T877A and F876L forms of AR, respectively. There was no significant effect on the growth of PC3 PCa cells which lack the AR. The effectiveness of these BF3 inhibitors was also confirmed when they were shown to reduce the endogenous expression levels of PSA in Enzalutamide-resistant cell lines. Even though the specific mechanism for Enzalutamide-resistance in MR49F cells is still unclear and may or may not be fully related to mutations in the AR, the effectiveness of compounds **9117**, **9114** and **9145** in these cells substantiates targeting an alternative binding site on the AR such as the BF3. The results

obtained from the ERα luciferase assay confirms that these inhibitors are specific to AR and do not effect ERα in human breast cancer cells.

In summary, the study resulted in development of novel class of anti-AR drugs chemotypes with an alternative mechanism of action which can overcome conventional anti-androgen resistance and exhibit strong antagonism in PCa cell lines. The only limitation of the study on benzoimidazole series was we did not achieve great potency that could potentially reflect *in vivo* applications. Nonetheless, these derivatives were promising enough to carry out further hit-to-lead optimization efforts.

Chapter 4: DEVELOPMENT OF INDOLE-BASED DERIVATIVES

4.1 Background

Although, the BF3 inhibitors reported[64] and discussed in chapter 3 demonstrated promising *in vitro* activity profile we did not achieve great potency that could potentially reflect *in vivo* applications. Hence we initiated the current study to identify another chemical class of compounds using *2-[(2-phenylethyl) sulfanyl]-1H-1,3-benzodiazole* (**9099**) [64] as template for further drug optimization.

4.2 Results

4.2.1 Identification of Indole Chemical Series by the Shape-Based Similarity Search

In previous chapter, we reported series of synthetic derivatives by substituting different functional groups on benzimidazole and benzene moieties which resulted in their increased anti-AR potency. However, derivatives with longer linker fragments connecting the aromatic systems did not generally demonstrate improved activity (with the exception of compound **9099)**. In the current study, we searched for chemicals containing various shorter linkers using compound **9099** as template. We employed Instant JChem, a 2D similarity searching tool from ChemAxon to search for structural analogues. Similarity search was performed against the ZINC database v12.0 containing ~18 million compounds.[144] All software parameters were set to their default values. The search resulted in 10 candidate structures. The identified compounds were mapped to the query template and ranked according to generated "Tanimoto coefficient (Tc)" values.

The identified molecules were inspected visually and only compound **13235** was selected because it contains a short linker that connects two aromatic systems and obtained Tc value above 0.8. Since compound **13235** represented a novel chemotype -3-[(E)-2-phenylethenyl]-1*H*-indole, it was used as a template for another similarity search and which led to the identification of a series of N-[1*H*-indol-3-ylmethylidene] aniline derivatives, compounds **13303-13309**. In the same way, each subsequent chemical series (3-[(E)-2-phenyldiazen-1-yl]-1*H*-indole, compounds **13226-13258**; 3-(3H-indol-2-yl)-1*H*-indole, compounds **13127-13221**; 3-(2, 3-dihydro-1*H*-indol-2-yl)-1*H*-indole, compounds **13163** and **13164**;) was identified based on the scaffold of active compounds from the previous search (compounds **13235**, **13303**, **13226** and **13127**) as shown in figure 4.1. Thus, from each similarity search steps, we obtained a different number of hits,

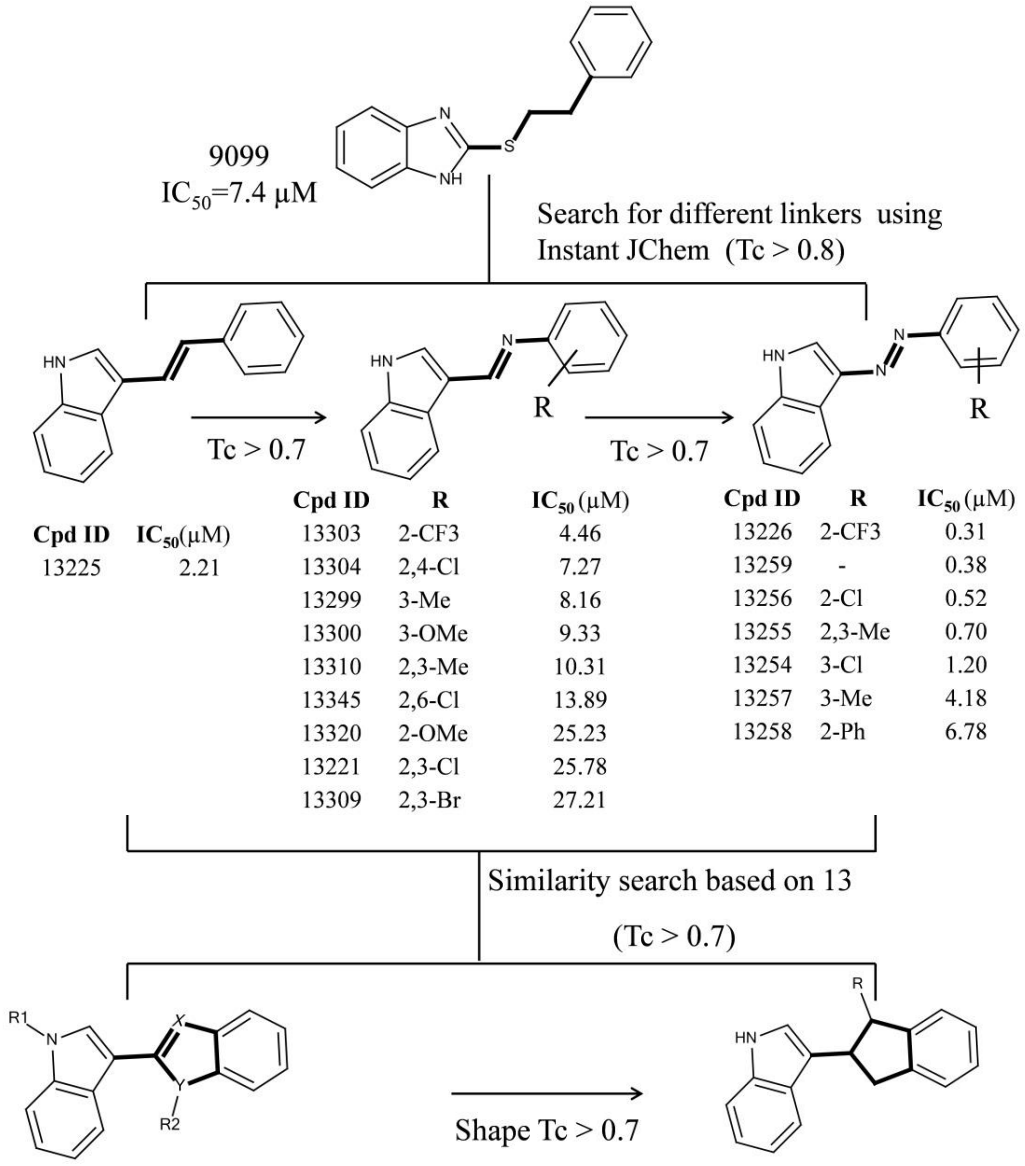
totaling 295 identified chemical structures. These compounds were further evaluated using our established *in silico* pipeline as discussed below.

4.2.2 Molecular Docking of Selected Compounds into the AR BF3 Pocket

The BF3 site represents a hydrophobic groove located adjacent to the AF2 pocket on the surface of the AR. Being a protein-protein interaction site, BF3 represents a challenging target. Nevertheless, it offers an attractive option for direct inhibition of the AR transactivation.

Using our *in-house* computational drug discovery pipeline, we virtually tested the selected 295 compounds. Our *in silico* pipeline included molecular docking, on-site rescoring, and consensus voting procedures (see Materials and Methods section for more details). Initially, all molecules were docked into the AR crystal structure (4HLW, 2.5Å resolution) using the Glide SP program [128]. Our previous studies indicated that charged amino acids Glu837 and Leu830 form an H-bond interaction and hydrophobic contacts with BF3 binders and are critical for protein-ligand coordination. Therefore, we have applied the corresponding H-bond and hydrophobic constraints during the docking. Compounds that received moderate-to-high score by Glide SP were selected and re-docked into the 4HLW structure using the eHiTS docking protocol.[129] In order to improve accuracy of the predicted binding orientation, the root-mean-square deviation (RMSD) was calculated for the docking poses generated by Glide and eHiTS programs. Only molecules with docking poses varying with RMSD < 2.0Å were subjected to further analysis.

At the next step, the selected ligands underwent additional on-site rescoring using the LigX program and the p*Ki* predicting modules implemented in MOE. With this information, a cumulative scoring of four different predicted parameters (Glide score, eHiTS score, RMSD, LigX score and pKi predicted by the MOE) was computed, with each molecule receiving a binary 1.0 score for every "top 20% appearance" (supplemental table 4.1). The final cumulative vote allowed for selecting 70 molecules associated with higher probability of being BF3 binders. These compounds were then visually inspected and 23 chemicals (compounds mentioned in figure 4.1) were selected.



Cpd ID	IC50(μM)
13225	2.21

Cpd ID	R	IC50 (μM)
13303	2-CF3	4.46
13304	2,4-Cl	7.27
13299	3-Me	8.16
13300	3-OMe	9.33
13310	2,3-Me	10.31
13345	2,6-Cl	13.89
13320	2-OMe	25.23
13221	2,3-Cl	25.78
13309	2,3-Br	27.21

Cpd ID	R	IC50 (μM)
13226	2-CF3	0.31
13259	-	0.38
13256	2-Cl	0.52
13255	2,3-Me	0.70
13254	3-Cl	1.20
13257	3-Me	4.18
13258	2-Ph	6.78

Cpd ID	Group	IC50(μM)
13127	X=CH, Y=N R1=H, R2=H	0.60
13166	X=CH, Y=N, R1R2=Me	1.69
13167	X=N, Y=CH, R1=H, R2=NO	1.91
13221	X=N, Y=N R1=H, R2=H	>25

Cpd ID	R	IC50(μM)
13163	H	0.31
13164	Me	1.20

Figure 4.1. The strategy used in order to obtain high active indole based AR BF3 inhibitors. The IC_{50} values shown for each compound are obtained from AR-eGFP assay.

To ensure that these compounds represent an entirely different chemical class of AR inhibitors, we compared them with previously reported BF3 binders [63, 64, 139] and with known anti-androgens (supplemental table 4.2) on the basis of structural and physicochemical similarity. The assembled set was clustered according to pairwise Tanimoto distances and using Daylight fingerprints. A clustering threshold of 0.5 resulted in 17 groups, highlighting structural diversity of the set (figure 4.2).

In addition, these structures were clustered according to their physicochemical similarity (figure 4.3). The heat map generated by matrix2png program[163] highlights the distances calculated in the first 3 principal components space (variance explained >95%) that originated from 9 drug-likeness descriptors (polar surface area, log*P*, log*S*, molecular weight, number of H-bond donors, H-bond acceptors, heavy atoms, rings, and rotatable bonds). From the resulting structural classification tree and the heat map, one can see that the identified chemicals are quite distinct from the previously reported BF3 binders and from conventional anti-androgens listed in supplemental table 4.2. Hence, these 23 chemicals were purchased and further investigated experimentally.

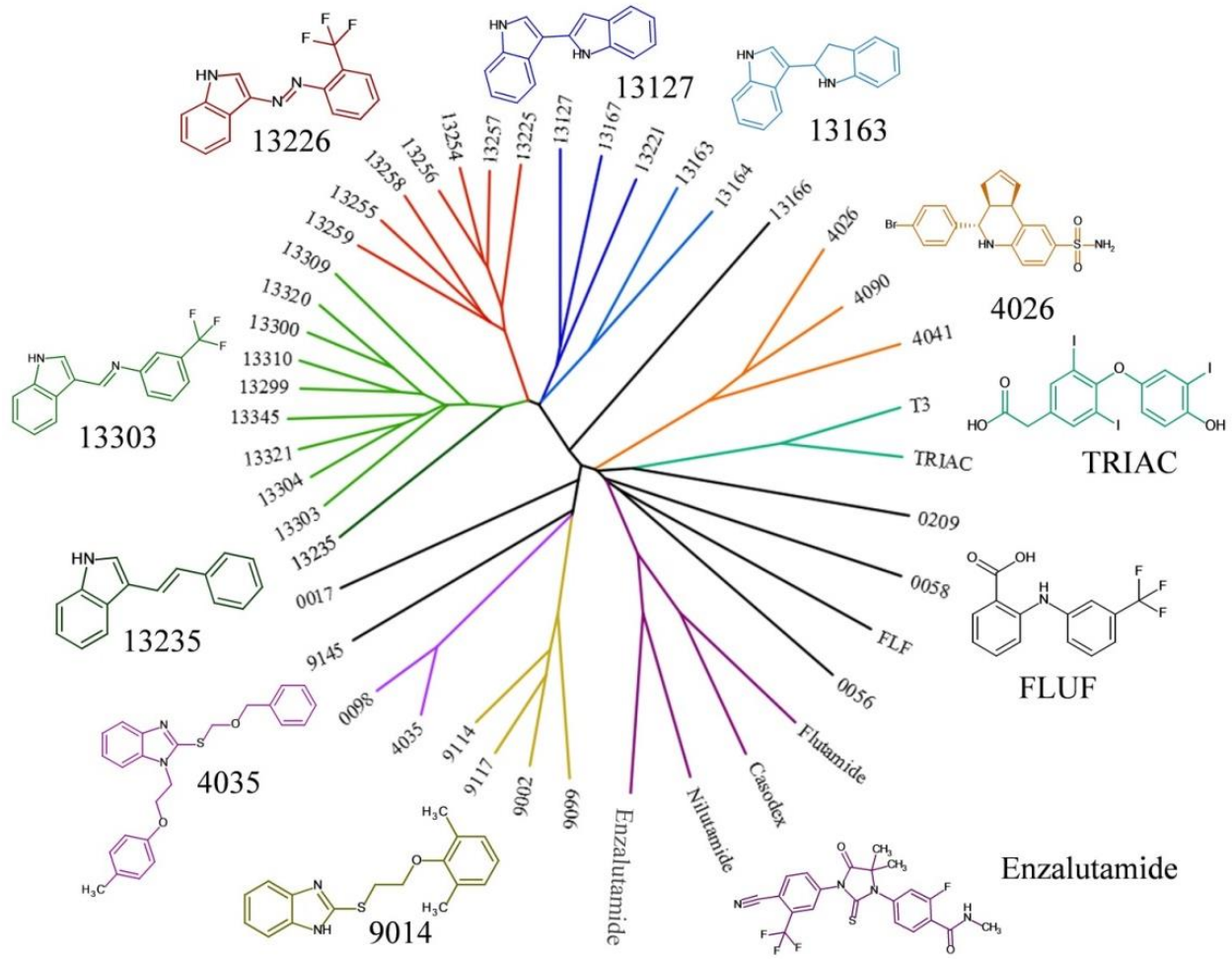


Figure 4.2. An un-rooted classification tree representing the structural diversity of reported BF3 inhibitors compared to published BF3 inhibitors and AR drugs. The tree is built based on pairwise Tanimoto distances between Daylight fingerprints of 44 AR inhibitors including 4 commercial anti-androgens and 17 previously published BF3 inhibitors. Only highest active inhibitor from each colored cluster is depicted.

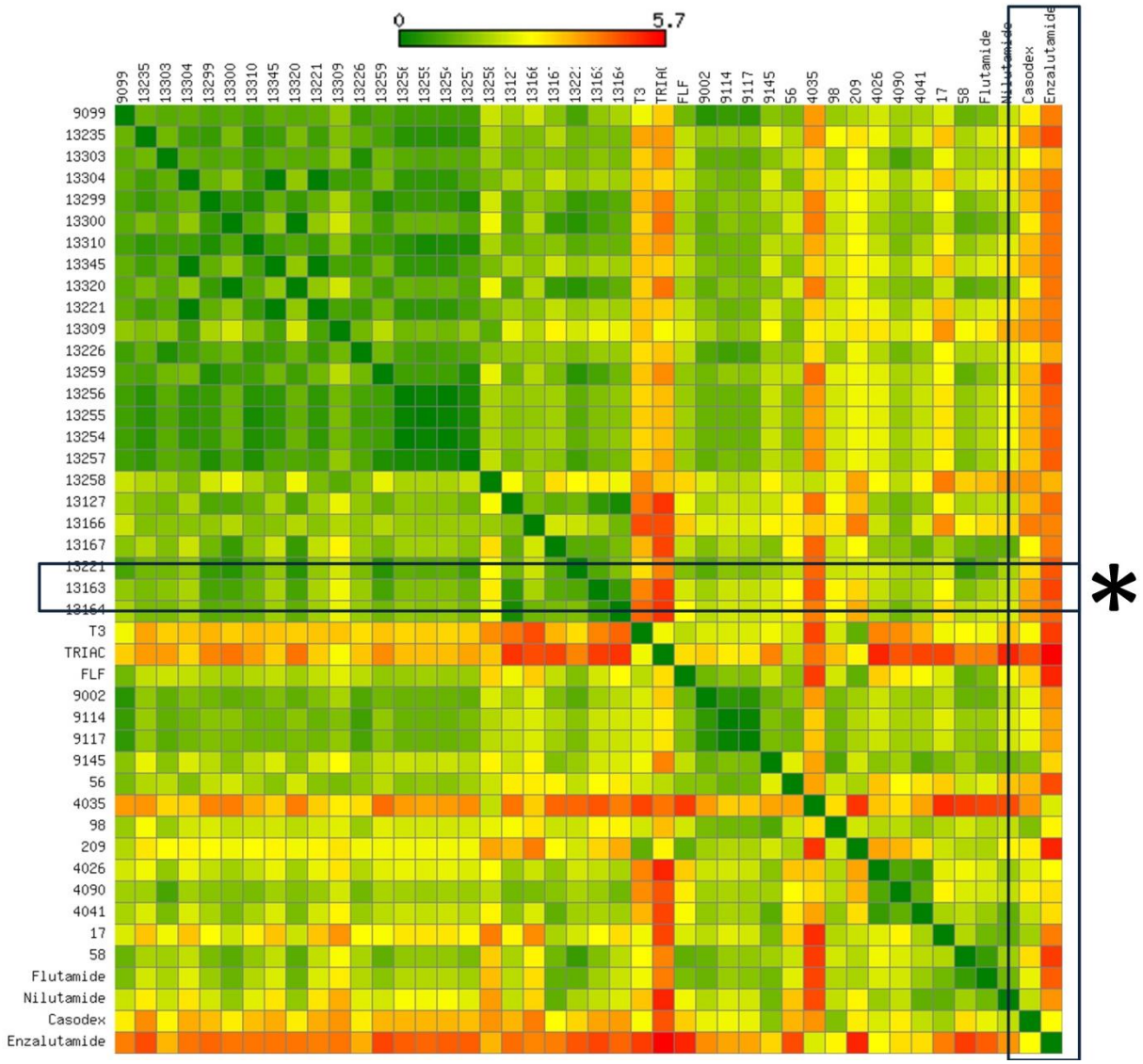


Figure 4.3. Comparative analysis of reported BF3 inhibitors and published/commercial AR inhibitors based on 9 physicochemical properties (polar surface area, logP, logS, molecular weight, number of H-bond donors, H-bond acceptors, heavy atoms, rings, and rotatable bonds). Green shades code for similarity whereas red shades code for dissimilarity. Compound **13163** has less or no similarity with previously reported BF3 compounds such as T3, TRIAC and ant-AR drugs Casodex and Enzalutamide (as highlighted by horizontal box). Asterisk (*) point highlights the difference between compound **13163** and Casodex, Enzalutamide.

4.2.3 Cell-Based Testing and *In Vitro* Characterization

All purchased compounds were screened for their ability to inhibit AR transcriptional activity using an eGFP AR transcriptional assay. [153] Since all the compounds exhibited >75% inhibition of AR transcription at the screening concentration of 3μM, they were all subjected to concentration-dependent titration along with Enzalutamide as a positive control. We established an IC_{50} value for Enzalutamide of 0.08 μM, whereas the tested compounds exhibited IC_{50} values in the range of 0.3-25 μM (figure 4.1).

In addition, we tested the most active derivative **13163** within the SRC2 peptide and androgen displacement assays to check if they displace the activator peptide from the AF2 site and androgen from the ligand binding pocket. It showed any significant activity in these assays at the concentrations measured, confirming that it likely target the BF3 pocket. Furthermore, biolayer interferometry (BLI) studies demonstrated a direct, reversible and dose-dependent interaction between the tested compound and the AR LBD. Figure 4.4A and B feature the eGFP IC_{50} curve and BLI graph for the highest active compound, **13163**.

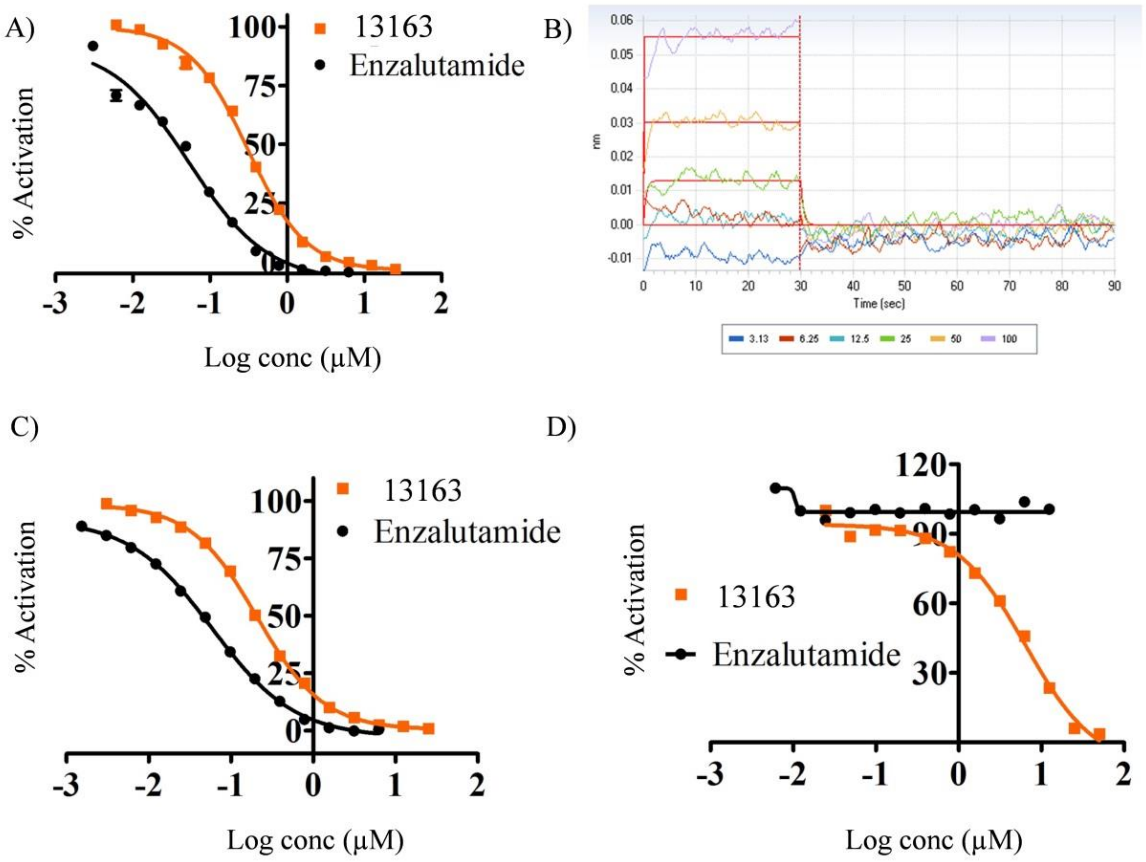


Figure 4.4. Activity profile of compound **13163**. A) Dose-response curve illustrating the inhibiting effect of compound **13163** (IC_{50} = 0.31 μM) and enzalutamide (IC_{50} = 0.08 μM) on the AR transcriptional activity in LNCaP cells. Data points represent the mean of two independent experiments performed in triplicate. Error bars represent the SEM for n = 6 values. Data were fitted using log of concentration (conc) of the inhibitors versus percent activation with GraphPad Prism 6. B) BLI dose-response curves (3–100 μM) reflecting the direct binding of **13163**. C) The effect of **13163** (IC_{50} = 0.21 μM) in comparison with enzalutamide on PSA (IC_{50} =0.09 μM) in a dose-response manner in LNCaP cells. D) The effect of **13163** in comparison with enzalutamide on PSA (IC_{50} = 6.02 μM) in a dose response manner in enzalutamide-resistant cells (MR49F).

4.2.4 Compound 13163 Stably Binds to BF3 Pocket

To determine the binding of **13163** at BF3 pocket, molecular dynamics (MD) simulations was performed for 50ns (see chapter 2. materials and methods). The stability of the receptor and ligand conformation was then assessed by the RMSD between the starting agonist conformation and the final conformation after 50ns MD simulations.

In figure 4.5A, the RMSD of **13163** is plotted in orange color and that of backbone heavy atoms of AR LBD is shown in grey color. The RMSD was calculated by using the initial complex conformation as the reference. The plateau of backbone RMSD over 50 ns indicates that the AR LBD was well equilibrated, and the **13163**'s average RMSD showing its value around 3.5Å over the maximum period (up to 40ns) clearly demonstrates that the ligand fits in the BF3 site well. We observed the RMSD shift at around 40ns and this is because there was the rotation of dihydroindole moiety in the 13163 ligand while its indole group is maintaining its pose close to initial confirmation.

In order to identify the representative binding pose of **13163**, we applied the cluster analysis to the **13163** ligand's conformations obtained from the 50ns MD production run. The clustering was conducted on the basis of RMSD of ligand's heavy atoms by using the hierarchical agglomerative clustering approach, and the five clusters were obtained. About 54% of the ligand conformations were assigned to the first ranked cluster, and its centroid conformation in the cluster (shown in yellow color in figure 4.5B) was found to be similar to the initial **13163** binding pose docked into the 4HLW crystal structure as depicted in green color in figure 4.5B.

To determine the representative binding pose of **13163**, we proceeded to the generalized born/surface area (GB/SA) model to calculate the binding free energy for all the protein-ligand complex conformations in the first ranked cluster, and defined the conformation that gave the lowest binding free energy as the representative binding pose. The binding free energy of the representative binding pose was calculated to be -30.6 kcal/mol, and we can see from the figure that the ligand is well surrounded by the residues in the BF3 site, although the BF3 site is the surface exposed binding pocket (figure 4.5C). Figure 4.5D demonstrates the residues in the BF3 site that had the frequent contact with the **13163** ligand during the MD simulations. We defined the contact between the ligand and residue when their closest distance between the constituent atoms is closer than 3 Å. For example, Phe673, Leu830, Asn833 and Glu837 had the contact with the ligand over 80% of total MD simulation time. This confirms that the intermolecular interactions are mainly driven by hydrophobic contacts. The H-bond interaction between Glu837 and **13163** was more prominent in docking studies. During MD studies, this interaction was observed approximately 10% of the total simulation time.

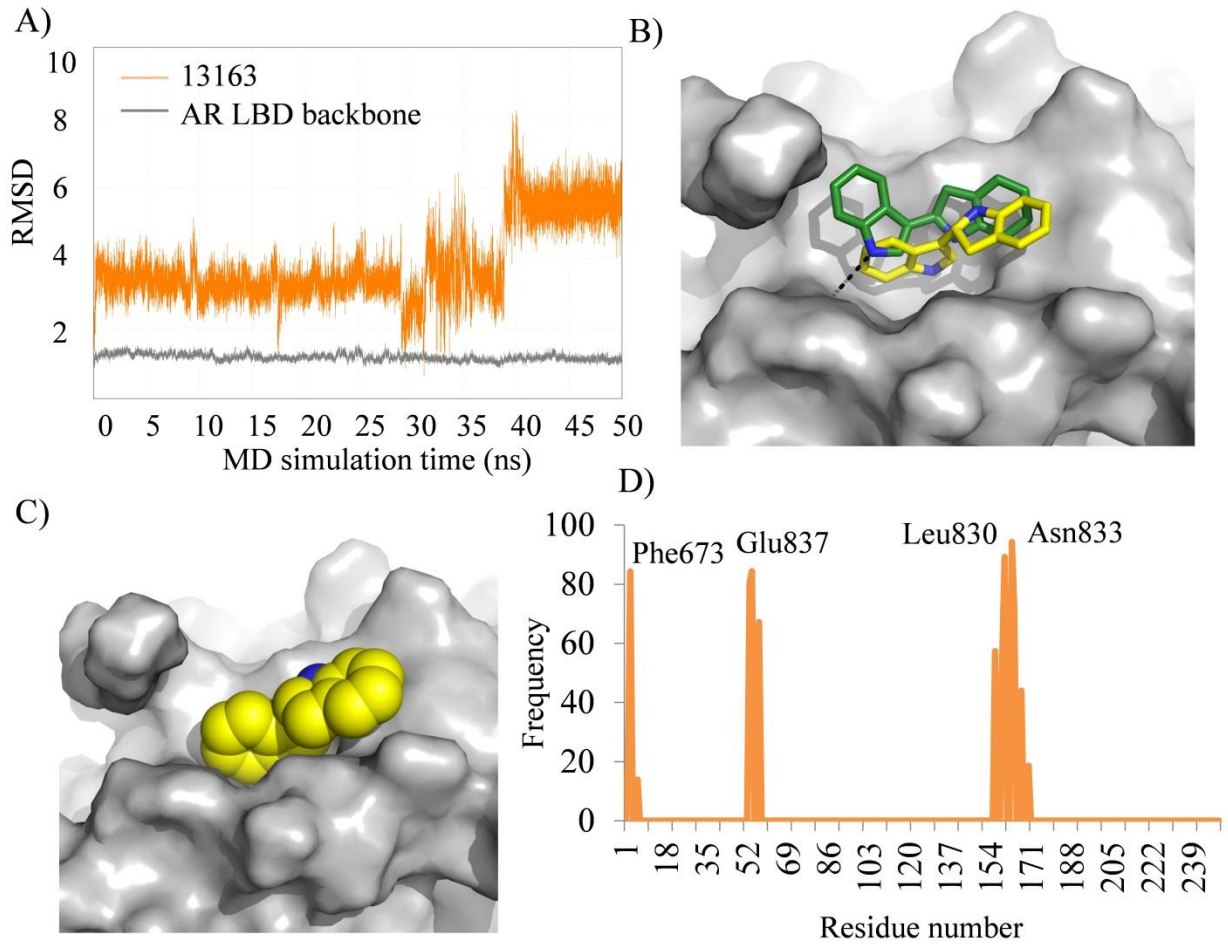


Figure 4.5. The analysis of molecular dynamics simulations performed on AR BF3-**13163** complex. A) The observed RMSD values observed for **13163** and AR LBD during 50ns MD simulations. B) Initial docked conformation predicted by Glide program (green) and representative docking pose and most stable conformation of **13163** obtained from MD simulations (yellow). C) Surface representation BF3 pocket and **13163**. It shows that the ligand occupies the BF3 site and bind tightly. D) List of BF3 residues that are in close contact with the ligand during MD simulations.

To confirm computational prediction of **13163**'s binding to the BF3 pocket, a mutagenesis study was performed on the AR. Preliminary results with mutants Phe673Glu, Glu837Ala (Figure 4.6A) and Asn833Trp (Figure 4.6B) confirmed that compound **13163** did not show any binding to the protein compared with the wild-type AR in a BLI assay. Because these residues are critical for protein-ligand interactions, compound **13163** binding has been abolished upon mutating them.

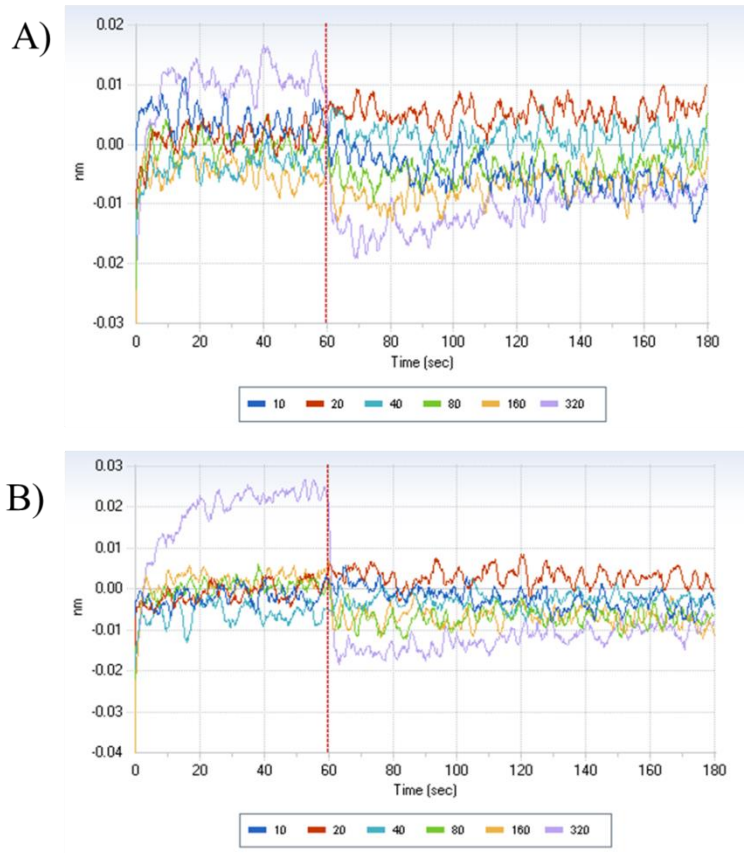


Figure 4.6. BLI dose-response curves (10-320 μM) showing that there is no binding of **13163** and mutant forms of AR BF3. A) Phe673Glu and Glu837Ala mutants. B) Asn833Trp mutant.

4.2.5 Compound 13163 Displaces Bag-1L Peptide from the AR BF3 Pocket

Recently, Jehle *et al* reported that Bag-1L protein interacts with AR BF3 pocket through its hexapeptide repeat sequence, GARRPR.[66] They demonstrated that one of our BF3 inhibitors mentioned in chapter 3 (**VPC-9114**) blocks the association between GARRPR motif and BF3 pocket. Therefore, to determine the binding of hexapeptide sequence at BF3 pocket, MD simulations was performed for 40ns. In figure 4.7A, the plateau of backbone RMSD over 40 ns (shown in grey) indicates that the AR LBD was well equilibrated, and the GARRPR motif's average RMSD (shown in orange) showing its value around 2.5Å throughout the simulation time indicates that the peptide fits well in the BF3 site. However, due to the presence of three highly flexible arginine residues in the hexapeptide sequence, RMSD value is increased when sidechains were considered in the calculation (blue curve).

Figures 4.7B and C illustrate binding mode of GARRPR motif and BF3 residues that are critical for the protein-protein associations. It can be clearly seen that the peptide makes a

network of hydrogen bond interactions with several BF3 residues. Phe673, Asn727, Glu829, Asn833, Glu837 and Arg840 had the contact with the peptide over 80% of total experimental time frame. During simulations, H-bond interaction between carbonyl oxygen of Arg_3 and sidechain of Arg840 was observed for more than 60% of the time. Similarly, the interaction between Arg_1 and sidechain of Asn833 was observed 79% of the total simulation time. The average bond distance of these interactions are estimated to be 2.6Å. The significance of these residues is confirmed by the mutagenesis study reported by Jehle *et al*. It has been observed that the interaction of Bag1L protein was strongly compromised when Asn833 and Arg840 were mutated compared to wild-type AR-Bag1L complex. Moreover, authors reported that the Proline of the GARRPR motif is a highly conserved residue among the peptides homologues to Bag1L. It has been observed in our MD simulations that Proline makes strong hydrophobic interactions with neighboring Phe673, Pro723 and Tyr734, hence, critical for the binding at AR LBD surface.

The superimposition of the binding orientations of **13163** and GARRPR motif inside the BF3 pocket as shown in figure 4.7D explains the mechanism behind the possible displacement of GARRPR motif from the binding pocket in the presence of a BF3 inhibitor. **13163** perfectly occupy the binding regions of hexapeptide motif and disrupt key interactions between the Bag1L protein and BF3 residues. According to MD simulations, the compound disrupts the H-bond interactions and van der Waals contacts formed between ARRP residues and BF3 pocket.

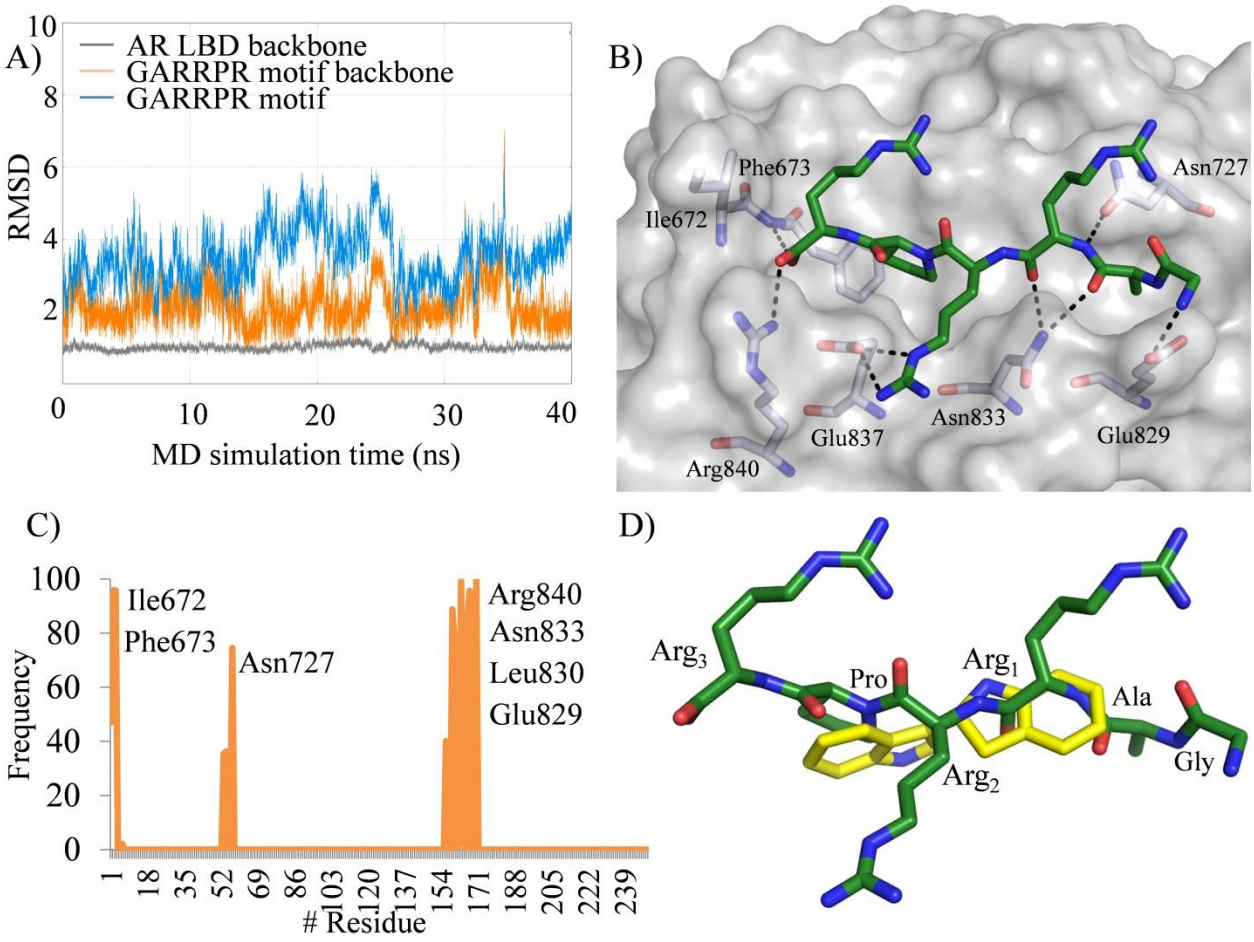


Figure 4.7. Predicted binding orientation of GARRPR motif of Bag1L protein inside the BF3 pocket. A) The observed RMSD values for GARRPR motif and AR LBD during 40ns MD simulations. B) Frequently observed binding conformation of the GARRPR motif during simulations. C) List of BF3 residues that are in close contact with GARRPR motif D) Superimposition of bound Bag1L peptide, GARRPR and **13163**.

Since **13163** has been developed as a selective AR BF3 inhibitor, we anticipated that the binding of this compound at the BF3 target should disrupt the interaction between GARRPR motif and AR. Therefore, the displacement of the bound peptide from the AR provides a direct validation for **13163**. Hence, the outcomes of the MD study were confirmed by a fluorescence polarization assay developed *in house* as described by Jehle *et al.*[66] We found that **13163** blocked the interaction of Bag-1L peptide with the AR with the corresponding IC_{50} value of 6.10 μM, whereas compound **13221**, characterized in prior experiments as non-AR disrupting and hence used as a negative control, exhibited no effect on peptide displacement (Figure 4.8A).

Interestingly, Casodex also demonstrated moderate displacement of Bag-1L from the BF3 pocket. This could be explained by Casodex binding to the androgen binding pocket of the AR, which then destabilizes the protein and allosterically affects the BF3. When Casodex was tested with AR-T877A mutant protein (where the drug acts as an agonist and stabilizes the AR) it did not displace Bag-1L peptide, whereas compound **13163** exhibited an unchanged potency against the mutant AR-T877A (IC_{50} value of 7.51 µM; see supplemental figure 4.1). Collectively, these observations strongly imply that compound **13163** binds directly to the AR BF3 site.

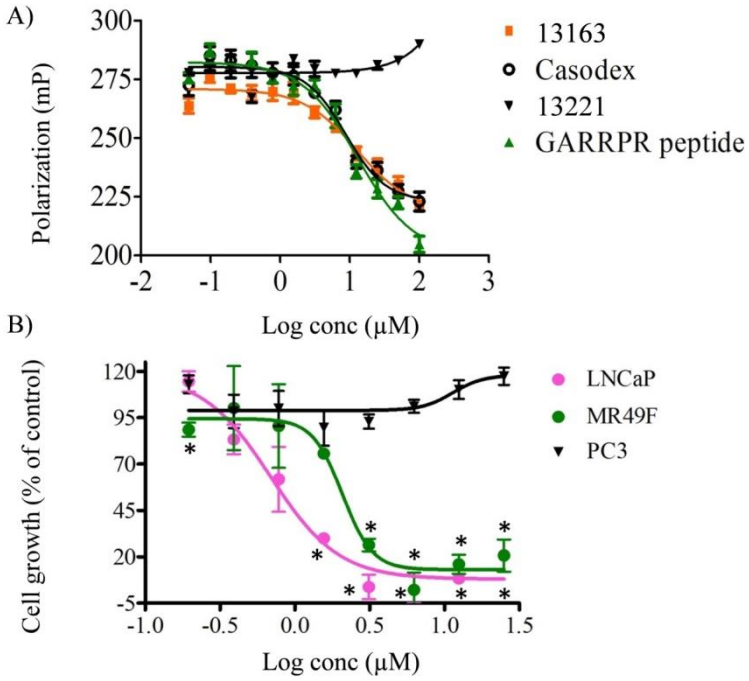


Figure 4.8. **13163** displaces Bag1L peptide and compound effect on PCa cell growth. A) Fluorescence polarization experiment showing competition between FITC-labeled Bag-1L (61-80) peptide (12.5nM) and serially diluted Casodex, **13163**, **13221**, unlabeled core GARRPR hexapaptide (100-0.05µM) for binding to AR-LBD (2 µM). The competition experiments were performed at conditions for maximum polarization of FITC-labeled peptide and AR-LBD. B) The effect of compounds **13163** on cell viability in LNCaP, Enzalutamide resistant cell line (MR49F) and PC3 cells. % cell viability is plotted in dose dependent manner. Data are presented as Mean ± SEM. A p value <0.05 was considered very significant effect (*) on LNCaP and MR49F compared with PC3 cells.

4.2.6 Activity Profile of 13163

Upon testing, **13163** induced a significant decrease in secreted PSA levels in LNCaP cells with the corresponding IC_{50} value established at 0.21 μM (figure 4.4C). Compound **13163** was further evaluated in Enzalutamide-resistant prostate cancer cell line. As anticipated, Enzalutamide was confirmed to be ineffective against MR49F line, whereas compound **13163** caused a significant reduction in the PSA (IC_{50} = 6.02 μM; see figure 4.4D). The anti-proliferative effect of **13163** was assessed in LNCaP, MR49F and PC3 cells by MTS assay. As shown in Figure 4.8B the compound is very effective in reducing the growth of both LNCaP and MR49F cells, achieving IC_{50} values of 0.71 and 2.01 μM, respectively. Moreover, compound **13163** did not show any effect on AR-independent PC3 cells, confirming its AR-specific activity.

The selectivity of a compound **13163** toward other nuclear receptors was evaluated as contract research by Life Technologies. Supplemental figure 4.2 demonstrates that compound **13163** does not have a significant effect on GR or ERα. The compound exhibited 80-100 fold decrease in inhibitor activity against ERα (IC_{50}=22 μM) and GR (IC_{50}=30 μM) compared to the AR (IC_{50}=0.31 μM). As anticipated, compound **13163** demonstrated a somewhat better, although ~7-fold weaker activity (IC_{50}=2.1 μM), against PR, as the AR and PR are more similar in their sequence in the BF3 region [164, 165] compared to the BF3 conservation between the AR and GR and ERα.

4.2.7 Compound 13163 Inhibits AR-dependent Growth of Xenograft Tumors *In Vivo*

The *in vivo* effect of **13163** was evaluated with both androgen-sensitive LNCaP and Enzalutamide-resistant MR49F xenografts. The IV, PO and IP serum profiles of compound **13163** (supplemental figure 4.3) suggest that it could be administered orally with substantial retention up to 24 h. The initial experiments demonstrated no systemic toxicity, and doses up to 200 mg/kg (twice a day) could be tolerated by the mice with no decrease in body weight monitored for 3 weeks.

A dose of 200 mg/kg of **13163** and 10mg/kg of Enzalutamide (control) administered twice a day was chosen based on the preliminary studies. The *in vivo* screening for tumor growth was initially done using the castration-resistant tumor xenografts model [166-169], in castrated hosts [154, 170]. When tumor regrowth was observed and the serum PSA rose to pre-castration levels, the

mice were treated with compound **13163** at 200 mg/kg. The growth of the tumor volume was effectively suppressed in this castration-resistant xenograft models with both LNCaP ($p < 0.01$, figure 4.9A) and MR49F cells ($p < 0.05$, figure 4.9C), compared to the vehicle control. Moreover, compound **13163** significantly decreased the serum PSA levels in the LNCaP xenograft model ($p < 0.01$, Figure 4.9B). However, there was no significant change in the PSA levels of MR49F xenograft bearing mice compared to the vehicle-treated mice.

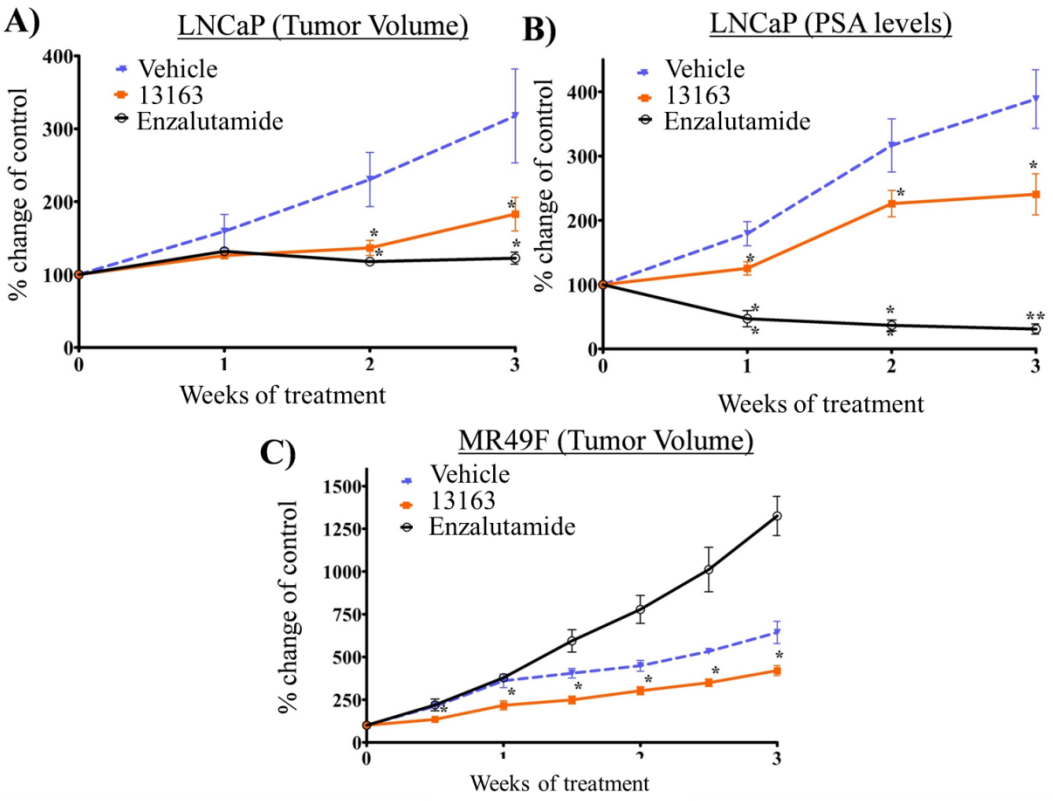


Figure 4.9. *In vivo* effect of compound **13163** in LNCaP and Enzalutamide resistant xenograft models. A) The *in vivo* effect of compound **13163** on the tumor volume of LNCaP mice xenograft. Data are presented as Mean ± SEM. A p value < 0.01 was considered very significant (*) ccompared to vehicle control. B) The *in vivo* effect of compound **13163** on PSA level of LNCaP mice xenograft. Data are presented as Mean ± SEM. A p value < 0.01 was considered very significant (*), A p value < 0.001 was considered extremely significant (**) compared to vehicle control. C) The *in vivo* effect of compound **13163** on the tumor volume of MR49F mice xenograft. Data are presented as Mean ± SEM. A p value < 0.05 was considered significant (*) compared to vehicle control.

4.3 Discussion

By harnessing the power of *in silico* modeling, combined with experimental evaluations, we identified a variety of BF3 binders which could effectively inhibit the AR at micromolar concentrations (as discussed in chapter **3**). Even though these compounds were effective in PCa cells *in vitro*, their potency was not sufficient for proper *in vivo* studies. Hence, the present work was initiated to develop a series of BF3 inhibitors that would be potent enough for evaluation in PCa human xenograft models and provide the foundation for subsequent therapeutic development.

At the onset, the previously identified BF3 binder - compound **9099** was selected as the chemical template for similarity search in the chemical space of a full ZINC database. As a result, 295 molecules were identified and subjected to docking in the AR BF3 site. Based on our knowledge of the BF3 residue-ligand interactions, compounds that interact with Glu837 and Leu830 were further selected, purchased and evaluated in our eGFP transcriptional assay[153]. Figure 4.1 illustrates that some of the selected chemicals, especially compound **13163,** demonstrate higher potency against the AR (IC_{50} = 0.31 μM) in the assay. Figure 4.4 illustrates that compound **13163** exhibits dose-dependent inhibition of AR and its direct binding to the AR LBD. In addition, we calculated a ligand-lipophilicity efficiency (LLE) score for 23 compounds reported in this study (see supplemental table 4.3). Since cell permeability, absorption, microsomal clearance and pharmacological clearance of the drugs are associated with *c*log*P*, it has been suggested that LLE is a critical factor to determine the quality of the lead candidates (LLE = pIC50-clogP)[171]. Compound **13163** received a high LLE score (LLE = 2.775) giving confidence to the premise that this chemical was a potential lead (see supplemental table 4.3).

The potency of **13163** was further confirmed by its ability to decrease endogenous PSA levels in PCa cells. The PSA inhibition value in LNCaP system (IC_{50} = 0.21 μM) was in agreement with the IC_{50} (= 0.31 μM) obtained from the AR eGFP assay, giving us confidence that the developed chemical is a true AR inhibitor. Compound **13163** also turned out to be more effective in reducing PSA levels in Enzalutamide-resistant MR49F cells (IC_{50} = 6.02 μM). Moreover, the compound possesses a strong anti-proliferative response on AR-dependent LNCaP and MR49F cells (IC_{50} of 0.71 and 6.02 μM), whereas there was no effect observed on AR-negative PC3 PCa cells. It should also be noted that LNCaP cells contain a mutated AR-T877A form of the receptor, while and MR49F contain an AR with the recently reported

mutation F876L that can turn Enzalutamide into agonist. [97] Since compound **13163** is a BF3-targeting AR inhibitor, we anticipated that this mutation would not affect its efficacy. As expected, it retained activity against both LNCaP and MR49F cell lines in all the experiments performed.

The MD simulation analysis performed on AR-13163 (figure 4.5) and AR-Bag1L (figure 4.7) complexes, separately, revealed the possible mechanism behind the displacement of GARRPR motif from the active site in the presence of a BF3 inhibitor. **13163** perfectly occupies the binding regions of hexapeptide motif and disrupt key interactions (H-bond interactions and van der Waals contacts) between the ARRP residues of Bag1L protein and BF3 residues. The analysis of inter-molecular interactions has provided valuable information and helped us in optimizing the structure of **13163** (discussed in further chapter).

These predictions of MD simulations were confirmed with a fluorescence polarization assay, where GARRPR portion of Bag-1L protein was used as a native BF3 ligand and was displaced by the compound in a concentration-dependent manner (IC_{50} = 6.10 μM). As the AR BF3 site is a recently established co-activator pocket [63] so far only two proteins (FKBP52[65] and Bag-1L[66]) have been reported to interact with it. The IC_{50} observed for compound **13163** in the Bag-1L peptide displacement assay is higher than its cellular activity (6.10μM Vs 0.31μM). It should be noted that this was observed in a cell-free assay. In reality, the BF3 site may accommodate a variety of proteins, many of which could contribute to or influence the cellular inhibitory activity of compound **13163**. Further studies are underway to determine the complete repertoire of interactions that can be blocked by the BF3 inhibitors.

Our *in vivo* studies demonstrated that compound **13163** achieves serum levels well above its IC_{50} and has a significant effect on tumor growth as well as serum PSA levels in the LNCaP xenograft model. However, the inhibition of serum PSA was minimal compared to the impact on tumor volume. This is likely because PSA production occurs independently of cell growth rate and may be influenced by various other factors, including hormonal and stromal milieu [172]. Upon testing the efficacy of compound **13163** in the MR49F Enzalutamide-resistant xenograft model, a substantial inhibitory effect on the volume was observed compared to the vehicle control, but with no significant effect on serum PSA levels were observed (data not shown). The difference in the efficacy of compound **13163** between LNCaP and MR49F tumor xenograft models might

be attributed to the mechanisms responsible for Enzalutamide resistance in MR49F, for example, expression of splice variants such as AR-V7 and/or factors not necessarily related to the AR.

In summary, we report the indole based compound, **13163** that exhibit very potent AR inhibition in androgen sensitive and Enzalutamide-resistant cells *in vitro* and *in vivo*.

Chapter 5: THE APPLICATION OF GLOBAL FREE ENERGY SCORING FUNCTION IN IDENTIFICATION OF DERIVATIVES OF VPC-13163

5.1 Background

As discussed in the previous chapters, the BF3 site is a shallow groove located on the surface of the AR. Therefore, compounds bind to this area (inhibitors discussed in the previous chapters) tend to exhibit poor/moderate correlation between their predicted binding energies (*i.e.* dock score) and biological activities (cellular IC_{50}s) as shown in supplemental figure 5.1. Hence, robust binary QSAR models were developed and applied as a scoring function to predict the binding affinity of synthetic derivatives of **13163** prior to their chemical synthesis. The entire study is discussed in part 1 section of this chapter. The developed approach helped us to obtain a better correlation between experimental activities and computational predictions in comparison with already existing scoring functions. More importantly, we developed a novel chemical scaffold, indole-quinoline, as a promising lead AR BF3 inhibitor. Part 2 of this chapter highlights the structure activity relationship and activity profile of this chemical series.

PART 1

5.1.1 Scoring Functions

Scoring functions are approximate computational methods widely used in structure-based drug discovery applications to predict the binding affinity (or free energy of binding) between molecules (commonly - protein and a ligand).[173, 174] During lead identification and optimization process, predicted binding affinity is used as a key parameter for compound selection and their advancement into the drug discovery pipeline. Free energies of binding are used to make assumptions about efficacy, selectivity and toxicity of drug candidates.[175-179] Moreover, free energy-based scoring function reduces the number of compounds that have to be screened in experimental assays. Unfortunately, the current docking and scoring techniques are trained to give the correct geometry, rather than correct free energy of interaction. [180, 181]

So far, a variety of scoring functions have been established, which can be roughly classified into three categories.[182] force field-based[183], knowledge-based,[184-186] and empirical[187, 188]. In order to provide better efficiency, these scoring functions are restricted and do not attempt to simulate certain physical processes that influence the process of binding. Therefore, scoring functions have limitations in their ability to rank-order and select small-molecules based on

predicted values. Although these scoring functions work slightly better for compounds that bind to deeper and buried binding pockets, the problem remains for molecules that occupy more exposed binding sites. One of efficient strategies to tackle such as problem is by combing ligand-based and structure-based approaches *i.e.* capturing protein-ligand interaction terms in a quantitative manner. Wikberg *et al* coined the term of *proteochemometrics* for such an approach.[189] Recently, Kramer and Gedeck summarized number of publications in their recent study that adopted similar approaches to derive models of free energy of binding.[148]

Previously, our group introduced various reactivity indices derived from the *linearity of free energy relationships* principle. All of these atomic and group parameters could be easily calculated from the fundamental properties of bound atoms and possess much defined physical meaning.[190-192] All of these parameters (also known as 'inductive' reactivity indices) have been expressed through the very basic and readily accessible parameters of bound atoms. their electronegativities (χ), covalent radii (R) and intramolecular distances (r). These descriptors were successfully used to predict the affinity of peptides. [193, 194]

Intermolecular distance plays a pivotal role while estimating binding affinities. Therefore, atomic pair potentials derived from interatomic distances can be expected to capture at least the distance preferences between protein and ligand to a reasonable extent. Hence, we developed a novel set of QSAR descriptors that captured interatomic distances between BF3 inhibitors and the binding pocket. First, the protein-ligand atom-pair distance dependent (APDD) descriptors were calculated based on general and specific categorization of element types. In addition, traditional 2D descriptors were calculated as implemented in MOE (Molecular Operating Environment) program. Robust binary QSAR models were developed and used as a scoring function to predict binding affinity of a new library of derivatives developed based on the structure of **13163**. Most promising synthetic derivatives were synthesized and evaluated experimentally.

5.2 Results

Surface exposed binding pockets or protein−protein interaction sites such as BF3 groove on AR are often considered as attractive opportunities for therapeutic targeting. However, identifying small molecules that modulate these sites is often difficult owing to issues such as

lack of a well-defined deep binding pocket. Therefore, predicting binding energy for BF3 inhibitor prior to their chemical synthesis and biological testing still remains a challenge. Hence, we proposed a novel QSAR like scoring function, built on a data set of known AR BF3 inhibitors. The distances between the general and specific atomic types of ligands and BF3 residues are encoded in the form of descriptors, which are subjected to a QSAR fitting procedure. It represents a proteo-chemo-metric approach to predict the binding free energy of novel synthetic derivatives of **13163**.

5.2.1 Development of Descriptors and QSAR Models

The sum of atomic distances between specific atomic type of BF3 residues and the docked ligands were captured as APDD descriptors (figure 2.1A). The atomic type was determined as 1) General.- based on geometry or hybridization and 2) Specific.- as per SMARTS atom typing scheme [147] (as explained in table 2.1; materials and methods). As shown in figure 2.1B three different metrics were used to calculate the distances. Once these descriptors were developed, QSAR models were generated using the WEKA software,[149] which is a collection of machine learning algorithms typically used in data mining studies. Anti-AR activity values measured by AR eGFP assay for all the BF3 inhibitors were transformed into binary form (1-active, 0-inactive).

First, several binary models were built using specific APDD descriptors. The robustness of these models was assessed by ROC, Accuracy, Specificity, Sensitivity and PPV values calculated as described in chapter 2 (materials and methods). These models received an average of ROC=0.826, PPV=0.799, Specificity=0.728, Sensitivity= 0.823 and Accuracy=0.781. These are shown in table 5.1. In addition, the quality of these models was assessed by root mean squared error (RMSE) with an average value of 0.493. The best model (ROC=0.876 and RMSE=0.278) was obtained with 10 descriptors, calculated using metric 3. The statistical difference in the model 1 and 9 confirms that precision of descriptor calculation (metric 1 versus metric 3) plays a critical role in determining the quality of the developed models.

In comparison, QSAR models were built using descriptors calculated based on atomic geometries described in MOE 2012. These models received an average of ROC=0.742, PPV=0.759, Specificity=0.657, Sensitivity = 0.753 and Accuracy=0.712 with an average RMSE value of 0.867 (supplementary table 5.1). The difference in the quality of models built using

specific APDD descriptors and general APDD descriptors is due to the specific categorization of each element. For example, specific classification represents a Carbon atom in 29 types whereas carbon has 3 types geometrically. Precise categorization of an element facilitates capturing of protein-ligand interactions in more specific manner. In addition, 2D descriptors were calculated for all 106 BF3 inhibitors to build QSAR models. These models received an average of ROC=0.774, PPV=0.750, Specificity=0.658, Sensitivity = 0.801 and Accuracy=0.738 with an average RMSE value of 0.812.

Table 5.1. The list of 35 QSAR models built based on distance-dependent descriptors classified as per SMARTS strings. Desc - Descriptors, PPV - positive prediction values.

Metrics	Desc	Model	ROC	PPV	Specificity	Sensitivity	Accuracy
3	10	Kstar	0.876	0.806	0.750	0.862	0.811
2	10	Kstar	0.874	0.806	0.750	0.862	0.811
3	20	Random Forest	0.859	0.839	0.767	0.825	0.802
2	10	Random Forest	0.856	0.839	0.778	0.852	0.821
3	10	Rotation Forest	0.856	0.774	0.714	0.842	0.783
2	20	Random Forest	0.843	0.839	0.744	0.776	0.764
2	20	KStar	0.843	0.806	0.733	0.820	0.783
3	20	KStar	0.843	0.806	0.733	0.820	0.783
3	10	KStar	0.839	0.823	0.761	0.850	0.811
2	15	KStar	0.837	0.823	0.761	0.850	0.811
1	10	Logit Boost	0.836	0.742	0.673	0.807	0.745
1	15	Rotation Forest	0.833	0.806	0.739	0.833	0.792
3	10	Random Forest	0.830	0.774	0.682	0.774	0.736
1	20	Rotation Forest	0.827	0.758	0.674	0.783	0.736
3	20	Rotation Forest	0.826	0.823	0.744	0.810	0.783
1	20	Bagging	0.823	0.806	0.739	0.833	0.792
1	10	Random Forest	0.822	0.823	0.738	0.797	0.774
3	15	Random Forest	0.820	0.806	0.714	0.781	0.755
2	20	Bayes Net	0.817	0.742	0.686	0.836	0.764
3	20	Bayes Net	0.817	0.742	0.686	0.836	0.764
2	10	Bayes Net	0.816	0.790	0.729	0.845	0.792
3	10	Bayes Net	0.816	0.790	0.729	0.845	0.792
2	15	Bayes Net	0.815	0.774	0.714	0.842	0.783
3	15	Bayes Net	0.815	0.774	0.714	0.842	0.783
1	10	Rotation Forest	0.812	0.823	0.750	0.823	0.792
2	10	Rotation Forest	0.810	0.774	0.702	0.814	0.764
1	15	Random Forest	0.810	0.806	0.727	0.806	0.774
2	10	Logit Boost	0.808	0.790	0.723	0.831	0.783
3	10	Logit Boost	0.808	0.790	0.723	0.831	0.783

Metrics	Desc	Model	ROC	PPV	Specificity	Sensitivity	Accuracy
1	15	ADTree	0.807	0.839	0.750	0.788	0.774
2	20	Logit Boost	0.807	0.726	0.653	0.789	0.726
2	10	Decision Table	0.805	0.823	0.750	0.823	0.792
3	10	Decision Table	0.805	0.823	0.750	0.823	0.792
1	15	Bagging	0.805	0.823	0.756	0.836	0.802
2	15	Decision Table	0.805	0.823	0.750	0.823	0.792
Average			0.826	0.799	0.728	0.823	0.781

5.2.2 The Validation of QSAR Models

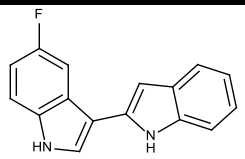
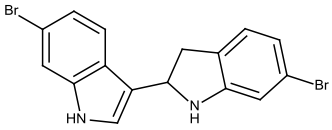
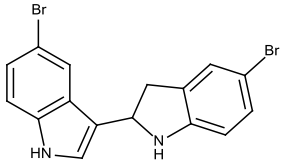
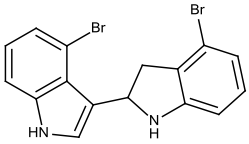
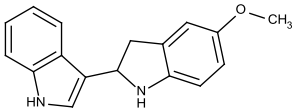
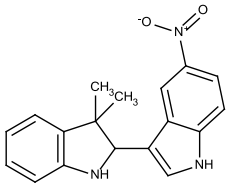
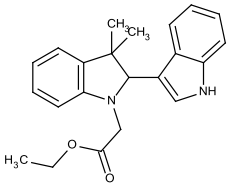
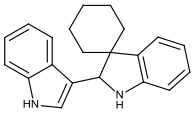
In order to validate the performance of the developed models, an external dataset of 400 derivatives was built as described in the chapter 2. Compared to models generated on general APDD descriptors and 2D descriptors, specific APDD descriptors received high ROC values *i.e.* in the range of 0.876 to 0.805 as shown in table 5.1. Moreover, they show significant sensitivity, selectivity and PPV and lower RMSE values, which determine the quality of QSAR models. Therefore, these models were used as a scoring function to predict the binding affinity of the derivatives.

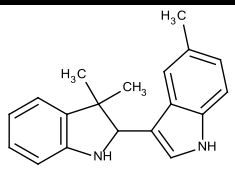
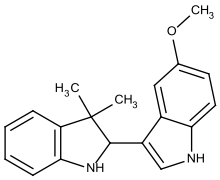
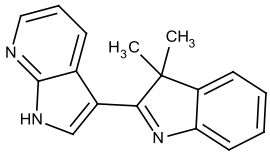
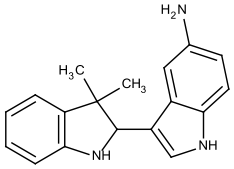
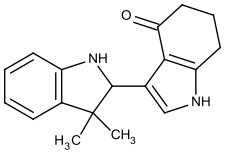
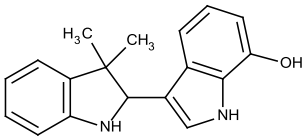
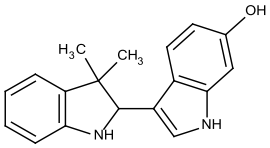
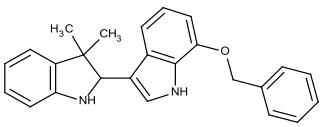
Preliminary Evaluation of Selected Derivatives. The range of descriptors selected in each model was compared to evaluate the applicability domain (AD) criteria for the developed models. This procedure estimated the reliability of predictions for the top-voted hits. Values of each descriptor for every top-voted chemical were checked to fit in the training descriptors value range. From this analysis we verified that all of the selected chemicals fit into suitable AD, with the percentage of reliable predictions being 100%.

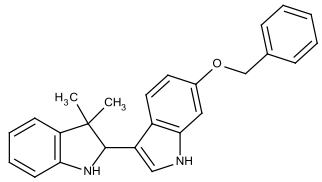
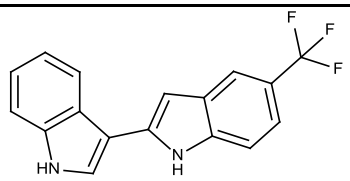
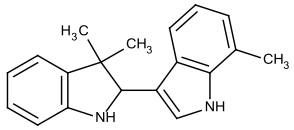
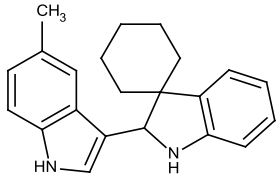
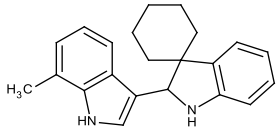
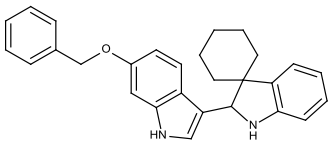
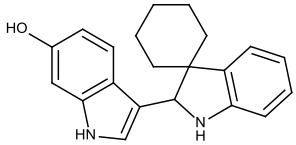
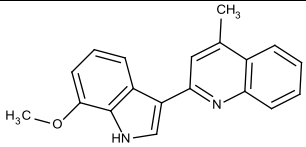
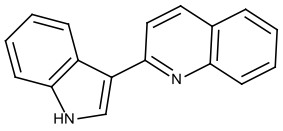
Consensus Analysis. The statistical results obtained from QSAR models indicate that different modeling techniques may have different advantages for predicting novel AR BF3 synthetic derivatives. Although the performances of our individual models are comparable, it is difficult to decide which model is the best one and which model should be chosen as a predictor for new derivatives. Thus it seems reasonable that the consensus approach can provide a better predictive ability than the individual models. The availability of several possible models, equally reliable for response prediction, highlights the need for methods able to preserve both model quality and diversity for model comparison. The strategy of majority voting (*i.e.* 1-actives, 0-inactives) is used to give predictions for all the developed derivatives. The final cumulative vote

with the maximum possible value of 35 is then used to rank the processed structures. Based on the synthetic feasibility, we selected 25 compounds for chemical synthesis. Activity profile and final score of QSAR predictions are shown in table 5.2.

Table 5.2. Activity profile of 25 synthetic derivatives. These compounds are selected and tested based on the developed on the predictions of QSAR models reported above.

VPC-ID	Structure	AR eGFP IC_{50} (µM)	PSA IC_{50} (µM)	Voting / 35 models
13500		1.21	1.84	26
13502		4.17	3.54	25
13503		4.52	6.36	25
13504		2.00	1.69	26
13509		2.61	3.12	26
13530		4.71	1.62	24
13532		2.83	3.14	25
13534		0.643	0.93	28

VPC-ID	Structure	AR eGFP IC_{50} (μM)	PSA IC_{50} (μM)	Voting / 35 models
13535		1.45	1.87	27
13536		2.17	2.52	27
13537		0.46	0.61	31
13538		7.7	Not tested	22
13539		Inactive	Not tested	13
13541		0.52	0.94	31
13542		5.71	Not tested	22
13543		13.69	13.19	18

VPC-ID	Structure	AR eGFP IC_{50} (μM)	PSA IC_{50} (μM)	Voting / 35 models
13544		11.39	Not tested	20
13545		4.85	2.92	25
13546		Inactive	Not tested	14
13548		3.21	4.24	27
13549		1.7	2.57	29
13550		11.75	Not tested	12
13551		1.12	1.4	29
13554		0.59	0.67	31
13562		0.160	0.21	33

5.2.3 Biological Evaluation of Synthetic Derivatives

The selected synthetic derivatives were tested and five of them demonstrated anti-AR activity and anti-PSA activity below 1μM (table 5.2). Importantly, compounds **13554** and **13562** belong to 2-(1H-indol-3-yl) quinolone chemical class demonstrated very potent at inhibition of AR transcriptional activity as measured by AR eGFP cellular assay. It should be noted that these derivatives received highest cumulative score by developed QSAR models (**13554=**31 and **13562=**33 out of 35 models).

5.2.4 Structure Based Prediction Versus QSAR Prediction

For comparing the performance of docking scoring functions with our QSAR prediction values we calculated enrichment factor (EF), which reflects the ability of the docking calculations to find actives throughout the background database compared to random selection. EF is defined as the ratio between the percentage of actives in the selected subset and the percentage in the entire database.

$$\text{EF subset} = \frac{\text{Hits selected/N subset}}{\text{Hits total/N total}}$$

First, the dock scores from different scoring functions such as Glide SP, eHiTS, Ligand explorer (LigX) and pKi are obtained for external set derivatives along with their consensus values from QSAR models. Enrichment values for 25 derivatives are summarized in table 5.3. LigX yielded slightly better enrichment value (at 60% dataset) compared to Glide SP, eHiTS and pKi. This could be due to LigX considers receptor and ligand flexibility while generating binding scores unlike other docking programs mentioned here. Importantly, superior enrichments are observed with QSAR consensus voting compared to docking scoring functions. On average, 100% and 92% of the active derivatives are found in the top 30% and 60% of the ranked data set, respectively, corresponding to better enrichment factors of 1.5 and 1.38 (table 5.3).

5.2.5 Significance of Consensus Analysis

Enthalpy (ΔH) and entropy terms (TΔS) are the key components while estimating binding energy (ΔG) between a protein and a ligand. Enthalpy reflects the strength of the protein-ligand interactions such as H-bond and VdW interactions whereas entropic contributions could arise from the solute (protein and the ligand), as well as from the solvent (usually water). Since BF3 site is surface exposed and a shallow pocket, it is necessary to calculate entropy terms accurately to obtain meaningful ΔG. Despite improvements over the last years, most scoring functions,

regardless whether they are empirical, force field based or knowledge based, still suffer from a rather poor correlation with experimental binding affinity.[195] This can be evidenced by ROC curves shown in figure 5.1. These ROC curves confirm that Glide SP, eHiTS, LigX and pKi scoring functions poorly ranked newly synthesized BF3 derivatives, based on their predicted score. However, consensus voting procedure based on specific APDD descriptors clearly outperformed traditional scoring functions. The consensus curve is on the top of all other models with the largest area of 0.89 for the prediction set. It also gives the highest sensitivity of 96% and specificity of 82.7%. The high sensitivity of 96% indicates that the consensus analysis can recognize more actives from a large number of chemicals. Furthermore, the similar values of the overall accuracy (89.8%) sets indicate the comparable external predictive abilities of the consensus analysis.

Table 5.3. Comparison of enrichment factor values calculated based on different scoring functions and QSAR voting pattern.

Scoring Function	% of the External Set		
Glide SP-Score	0.86	1.04	0.87
eHiTS-Score	0.64	0.81	0.95
LigX	0.86	1.15	1.03
dock_pKi	0.86	0.92	0.95
QSAR voting	1.50	1.38	1.11

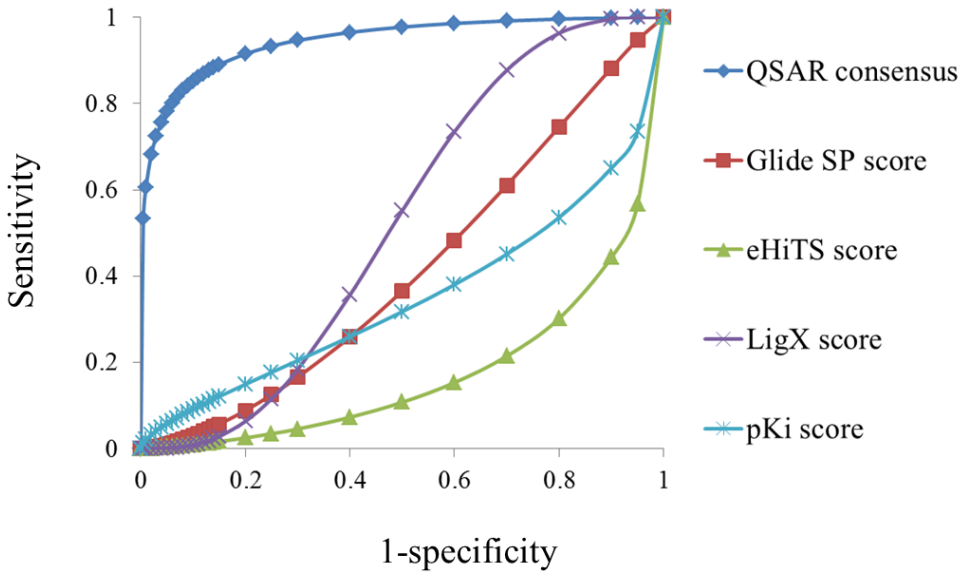


Figure 5.1. Comparison of ROC curves generated by QSAR models versus Glide, eHiTS, LigX and pKi docking programs. The curve generated by consensus voting of QSAR models (ROC value of 0.89) clearly outperformed other scoring functions.

5.2.6 Interpretation of the Descriptors

The most important and frequently occurred specific APDD descriptors are listed in table 5.4. These descriptors are straightforward to interpret atoms belonging to residues Ile672, Val676, Pro723, Glu829, Asn833 and Glu837 are critical for protein-ligand interactions. Most importantly, carbonyl oxygen of Glu837 side chain is critical for the binding and activity of BF3 inhibitors since it forms a strong H-bond with NH group of indole moiety (GLU837-O12.sp2 # H11.sp3). In order to test the significance of 'GLU829.O12.sp3 # 0_O2.sp' descriptor, we designed a derivative, **13541** with OH group at 7-indole position. As anticipated, carbonyl oxygen of Glu829 contributes positively for the prediction since it interacts with hydroxyl group of **13541**. The activity of compound **13542** decreased 10-times compared to **13541** since OH-group is located at 6-indole instead of 7th position. The replacement of dihydro-indole group with quinoline moiety favored the activity. The scaffold replacement resulted in identification of compound **13562** with an IC_{50} of 0.16µM.

The number of protein aliphatic hydrogens of Ile672, Val676, Pro723 and Asn833 forms other most important descriptors. They stand for hydrophobic rich feature of the binding pocket. Also, the number of halogens attached aliphatic/aromatic carbons of compounds and protein hydrogens connected to aliphatic carbons are critical.

Table 5.4. The list of most important and frequently occurred specific APDD descriptors.

Descriptor (protein atom # ligand atom)	Occurrence (out of 35)	Meaning
GLU837-O12.sp2 # H11.sp3	27	The distance between Glu837 carbonyl oxygen and Hydrogens attached to amine group
GLU829.O12.sp3#0_O2.sp#0	20	The distance between Glu829 carbonyl oxygen and Oxygen atom of the hydroxyl group ligands.
PRO723.H1.sp3#0_C4.sp3#0	21	The distance between Pro723 aliphatic hydrogens and all carbons bearing Halogens
ILE672.H1.sp3#0_H1.sp3#0	29	The distance between aliphatic hydrogens of Ile672 and aliphatic hydrogens of ligands
LEU830.H1.sp3#0_C4.sp3#0	29	The distance between Leu830 aliphatic hydrogens and all carbon bearing Halogens
PRO723.H1.sp3#0_C7.sp#0	28	The distance between Pro723 aliphatic hydrogens and Carbon of ligands' CN group
ASN833.C1.sp3#0_C1.sp2#0	29	The distance between aliphatic hydrogens of Asn833 and aliphatic hydrogens of ligands
VAL676.H1.sp3#0_C21.sp2#0	26	The distance between aliphatic hydrogens of Val676 aromatic carbon attached to aliphatic C of ligands

In summary, we developed a customized scoring function to predict binding free energies that have been trained on *in house* developed AR BF3 inhibitors. The scoring function is fit in a QSAR-like manner with descriptors coding for specific interactions between AR BF3 pocket and its binders. The developed APDD descriptors are straightforward to interpret and have never before been used in drug design. In its current state, the scoring function presented is useful for ranking derivatives and helps in prioritizing compounds for chemical synthesis. An important outcome of this study was that we identified a novel chemical scaffold, indole-quinoline, as a promising lead AR BF3 inhibitor, which is studied in detail in part 2.

PART 2

5.2.7 Rational Design, Synthesis, and Evaluation of Indole-quinoline Series

As discussed previously, we identified compound 13562 that belong to indole-quinoline series with an IC_{50} of 0.16μM. Hence, we initiated a lead optimization based on the observations from 1) molecular dynamic simulations performed on AR BF3-**13163** complex and AR-Bag1L

complexes as discussed in sections **4.2.4** and **4.2.5** 2) interpretation of APDD descriptors listed in table 5.4.

Compound **13562** forms a hydrogen bond between the NH indole moiety and side chain carboxyl of the Glu837. The importance of this interaction was evaluated by replacing NH with a methyl group. As molecular docking studies predicted, it resulted in 10-fold decrease in the activity (**13571** = 1.20 μM). Next, we focused on testing the effect of substitutions around the indole ring of **13562**. Evaluation of derivatives **13566**, **13569**, **13622** and **13582** revealed that groups with different van der Waals radius such as F, Cl, Me and Br affect the potency. For example, **13569** with F at 7-indole position exhibited an IC_{50} of 1.60μM whereas **13566** and **13622** with methyl and Cl, respectively, showed 20-fold increment in the activity compared to **13569** (**13566**=0.080μM and **13622**=0.10μM). Similarly, **13582** harbors Br at 7-indole position showed improved activity (IC_{50}=0.052 μM). Since methyl and Br were favorable for the activity, we designed compounds with same groups at 5-indole and 6-indole positions. These modifications hampered the activity by 20 to 30 fold compared to **13566** confirming that methyl and Br are favored only at 7-indole position. In order to study the significance of methyl group at 7^{th} position we initiated a molecular dynamic study which is discussed in further section.

An effort was also made to design derivatives by replacing quinoline with different rings. However, these modifications did not yield any potent derivatives except for **13603** (IC_{50}=0.63 μM).

5.2.8 VPC-13566 Stably Binds to BF3 Pocket in MD simulations

The binding pose of compound **13566** predicted by Glide SP program is used a starting point for the 100ns-MD simulations. The stability of the receptor and ligand conformation was then assessed by the RMSD between the starting agonist conformation and the each snapshot during MD simulations, and GB/SA binding free energy calculation.

In figure 5.2A, the RMSD of **13566** is plotted in orange color and that of backbone heavy atoms of AR LBD is in grey color. The RMSD was calculated by using the initial complex conformation as the reference. The plateau of backbone RMSD over 100ns indicates that the AR LBD was well equilibrated and the **13566**'s average RMSD showing its value around 2Å over the entire 100ns clearly demonstrates that the ligand fits in the BF3 site very well with its binding orientation similar to its initial conformation. We observed the RMSD shift at around 67

ns and this is because there was the rotation of quinoline group in the **13566** ligand while its 7-methylindole group is keeping its binding pose.

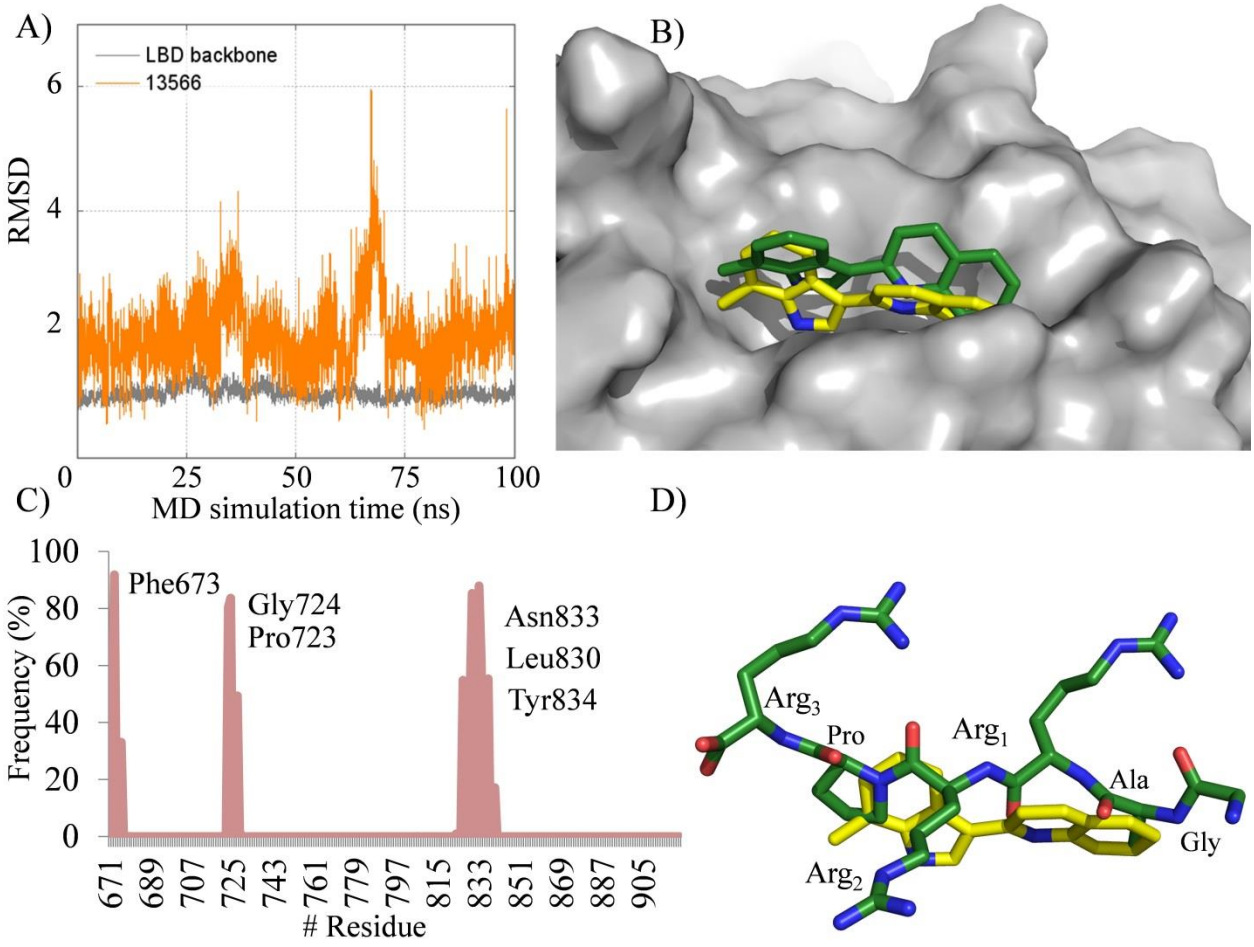


Figure 5.2. The analysis of molecular dynamics simulations performed on AR BF3-**13566** complex. A) The observed RMSD values observed for **13566** and AR LBD during 100ns MD simulations. B) Initial docked conformation (green) and representative docking pose of **13566** obtained from MD simulations (yellow). C) List of BF3 residues that are in close contact with the ligand during MD simulations D) Superimposition of bound Bag1L peptide, GARRPR and **13566**.

In order to identify the frequently occurring binding pose of **13566** during the MD simulations, we applied the cluster analysis to the **13566** ligand's conformations obtained from the total 100ns production MD run. The clustering was conducted on the basis of RMSD of ligand's heavy atoms by using the hierarchical agglomerative clustering approach, and the five

clusters were obtained. About 45% of the ligand conformations were assigned to the first ranked cluster, and its centroid conformation in the cluster was found to be similar to its input conformation (shown in yellow in figure 5.2B). To determine the most stable binding pose, we proceeded to the GB/SA binding free energy calculation for all the protein-ligand complex conformations in the cluster and defined the conformation that gave the most stable binding free energy. The binding free energy of the representative binding pose was calculated to be -33.6kcal/mol. Figure 5.2C shows the residues in the BF3 site that had the frequent contact with the **13566** ligand during the MD simulations. For example, Phe673, Leu830 and Asn833 had contact with **13566** over 80% of total MD simulation time.

Earlier analysis of AR BF3-**13163** complex revealed the mechanism behind the possible displacement of GARRPR motif from the BF3 site. The Pro residue of GARRPR motif makes strong hydrophobic interactions with the BF3 residues. Due to the presence of 7-methyl indole moiety and quinoline in **13566**, the compound binds more tightly at the binding pocket and disrupts hydrophobic interactions between Pro/Ala and BF3 residues (figure 5.2D). Hence, greater free energy of binding for **13566** was observed compared to **13163**. This explains why **13566** exhibits stronger inhibition of AR compared to **13163** (0.31 μM versus 0.80 μM).

5.2.9 Activity Profile of VPC-13566

As shown in figure 5.3 the compound resulted in IC_{50} values of 0.08μM and 0.09 μM for its anti-AR and anti-PSA activity in LNCaP cells and exhibited direct reversible interaction with AR LBD. Also, it caused a significant reduction in the PSA expressed in Enzalutamide-resistant MR49F cells (IC_{50} = 1.7μM). To determine the efficacy of **13566** on the viability of various PCa cell lines, its activity was assessed in LNCaP, MR49F and PC3 cells. The results indicated that **13566** was very effective in inhibiting the growth of both LNCaP and Enzalutamide-resistant cells, achieving IC_{50} values of 0.11 and 0.51μM, respectively. Moreover, 13566 did not show any effect on AR-independent PC3 cells, confirming its AR-specific action (figure 5.4A). Furthermore, **13566** did not affect the expression levels of AR in PC3 cells transfected with wild-type AR (figure 5.4B). To further confirm **13566** is a true BF3 binder we tested the ability of increasing concentrations of **13566** to displace a FITC labelled Bag1L peptide from the BF3 pocket of a purified LBD using a TR-FRET assay. Indeed **13566** was able to displace Bag1L peptide with a dissociation constant (K_d) of 19.2μM. However the compound **14449**,[88] a reported inhibitor that targets the DNA Binding domain of AR was not able to displace Bag1L from its

pocket (figure 5.4C). Overall, these results confirm that **13566** is a potent and true AR BF3 inhibitor.

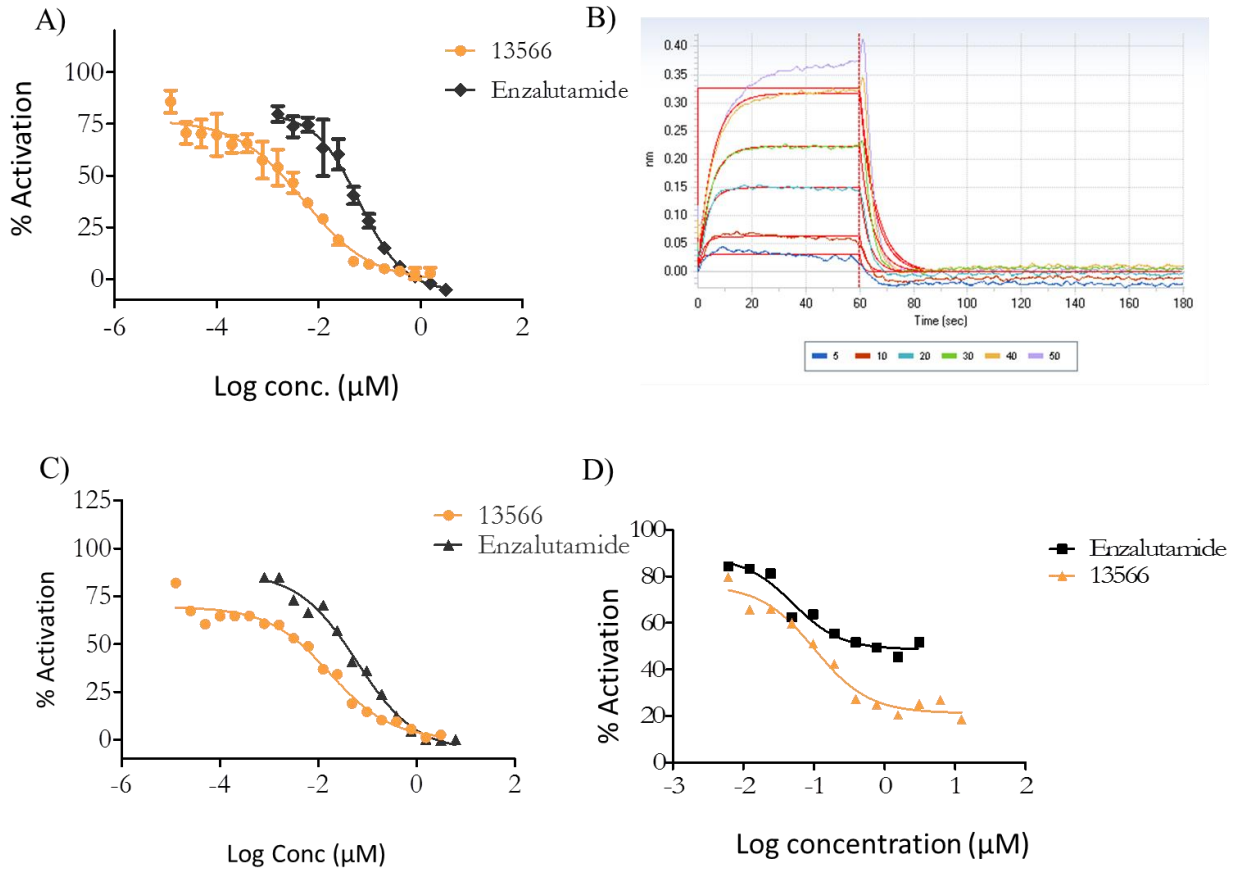


Figure 5.3. Activity profile of **13566**. A) Dose-response curve illustrating the inhibiting effect of the **13566** and Enzalutamide on the AR transcriptional activity in LNCaP cells. Data points represent the mean of two independent experiments performed in triplicate. Error bars represent the standard error of the mean (SEM) for n = 6 values. Data was fitted using log of concentration of the inhibitors vs % activation with GraphPad Prism 6. B) BLI dose-response curves (3-100µM) reflecting the direct binding of the **13566**. C) Effect of **13566** in comparison to Enzalutamide on PSA in dose response manner in LNCaP cells D) Effect of **13566** in comparison to Enzalutamide on PSA in dose response manner in Enzalutamide resistant cells.

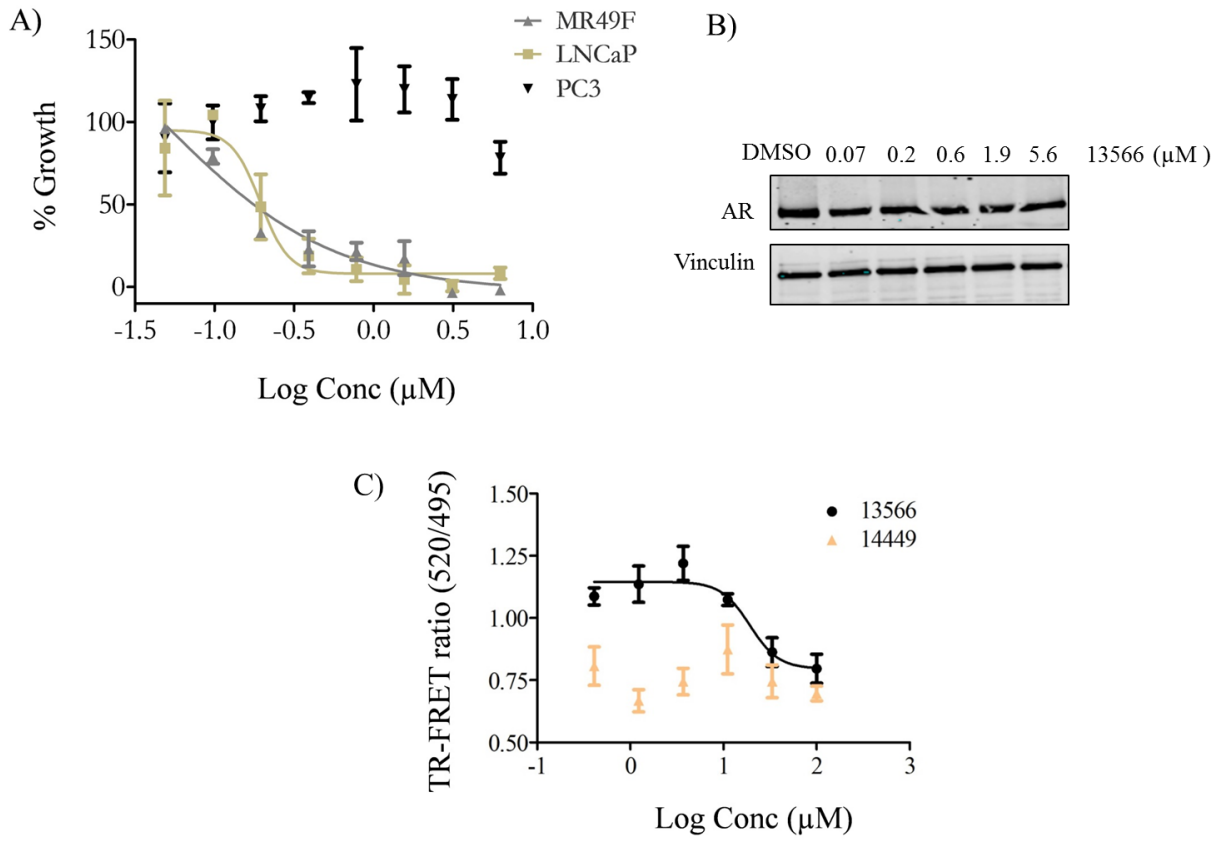


Figure 5.4. Further characterization of **13566**. A) The effect of **13566** on cell viability in LNCaP, MR49F and PC3 cells. % cell viability is plotted in dose dependent manner. Data are presented as Mean ± SEM. A *p* value <0.05 was considered very significant effect (*) on LNCaP and MR49F compared with PC3 cells. B) The effect of **13566** on the expression levels of AR in PC3 cells transfected with wild-type AR. C) Displacement of Bag1L peptide from BF3 pocket as measured by TR-FRET assay. Compound **14449**, used as a negative control did not show any effect since it does not bind to BF3 site.

5.2.10 VPC-13566 Inhibits AR-dependent Growth of Xenograft Tumors *In Vivo*

The *in vivo* effect of **13566** was evaluated with both androgen-sensitive LNCaP and Enzalutamide-resistant MR49F xenografts. The initial toxicity experiments demonstrated no systemic toxicity and doses up to and including 100 mg/kg twice a day could be tolerated by the mice with no decrease in body weight for 3 weeks.

A dose of 100 mg/kg administered twice a day was chosen based on these preliminary studies. The *in vivo* screening for tumor growth was initially done using the castration-resistant tumor xenografts model [166-169], in castrated hosts [154, 170]. When tumor regrowth was observed

and the serum PSA raised to pre-castration levels, the mice were treated with **13566** at 100 mg/kg. The growth of the tumor volume was effectively suppressed in this castration-resistant xenograft models with LNCaP ($p < 0.01$, figure 5.5A) compared to the vehicle control. Moreover, **13566** significantly decreased the serum PSA levels in the LNCaP xenograft model ($p < 0.01$, Figure 5.5B). The effect of **13566** in MR49F xenograft models is currently under investigation.

These results clearly indicate that **13566** could effectively inhibit androgen sensitive LNCaP xenograft growth *in vivo*, suggesting that this class of AR inhibitors has the potential to yield an AR targeting drug which could also be useful in the treatment of patients with castration-resistant tumors.

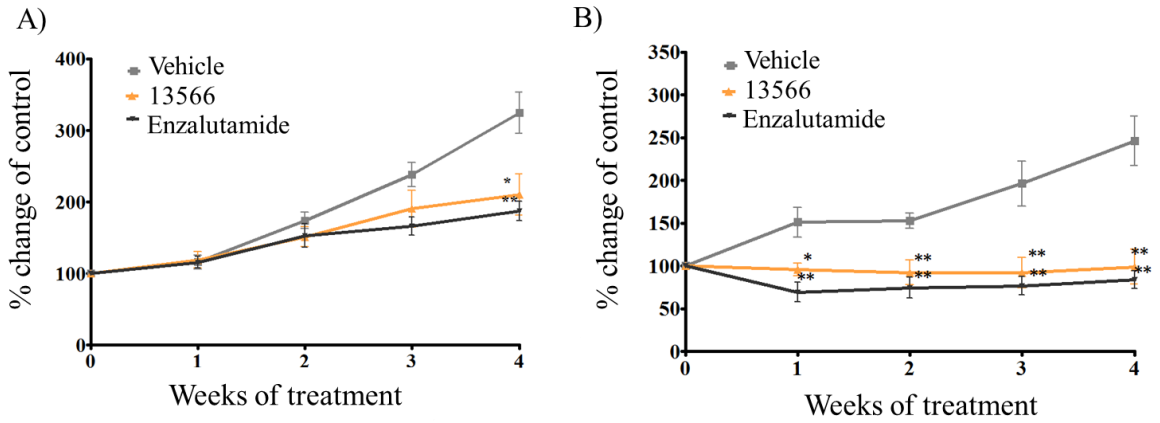


Figure 5.5. *In vivo* effect of **13566** in LNCaP xenograft model. A) The *in vivo* effect of **13566** on the tumor volume. Data are presented as Mean ± SEM. A p value < 0.01 was considered very significant (*) ccompared to vehicle control. B) The *in vivo* effect of **13566** on PSA level. Data are presented as Mean ± SEM. A p value < 0.01 was considered very significant (*), A p value < 0.001 was considered extremely significant (**) compared to vehicle control.

5.3 Discussion

In this chapter, we report the development of a novel scoring function for molecular modeling. This approach was developed in order to obtain a good correlation between *in silico* prediction and experimental activities of the studied series of AR BF3 inhibitors. The customised scoring function fits in a QSAR-like manner with descriptors coding for specific interactions between AR BF3 pocket and its binders. Compared to the traditional, universal scoring functions

(such as Glide and eHiTS), our customized approach demonstrated superior performance *i.e.* higher ROC values in terms of predicting new synthetic derivatives. An important outcome of this study was that we identified a novel chemical scaffold, indole-quinoline as a promising lead AR BF3 inhibitor.

The computer-aided optimization of indole-quinoline chemical class resulted in the development of the most potent BF3 inhibitor, **13566**. The activity profile of this confirmed that this compound has the potency comparable to Enzalutamide - a recently approved FDA drug for prostate cancer. In *in vitro* experiments **13566** demonstrated ant-AR activity. Importantly, the compound reduced the growth of LNCaP and Enzalutamide-resistant MR49F cells which contain clinically relevant mutant forms of AR, T877A and F876L, respectively. Further *in vivo* characterization of **13566** confirmed that the compound reduced PSA levels and shrinks PCa tumors in LNCaP mice models. A major limitation of 13566 was that it exhibited moderate microsomal stability ($T_{1/2}$ = 30 mins). Therefore, further lead optimization is ongoing to improve the physicochemical properties, and acquire a better microsomal stability and pharmacokinetics for preclinical studies.

There were no preceding investigations on targeting the AR BF3 site or other nuclear receptors by potent inhibitors. Our work provides the first proof-of-concept for targeting the AR BF3 by small-molecule inhibitors for the treatment of advanced PCa, which promises a new field of developing therapeutics utilizing AR BF3 as a target. More importantly, this work also demonstrated the BF3 is potentially a druggable site, and provided a new avenue of targeting nuclear receptors through a different mechanism of action, which is less likely to be associated with drug resistance.

Chapter 6: CONCLUSIONS

6.1 Summary of the Study

This thesis describes the development of novel small molecule inhibitors that selectively target BF3 regulatory site of human androgen receptor and can serve as prototypical therapeutics for the treatment of advanced prostate cancer (PCa).

PCa represents a common malignancy and second leading cause of male cancer-related deaths in North America. Since the discovery of the androgen dependence of prostate cancer in 1941 by Huggins and colleagues,[196] androgen deprivation therapy has remained the mainstay of PCa treatment.

Notably, all marketed anti-androgens including the latest Enzalutamide share similar chemical scaffold and exhibit similar binding mode of action *i.e.* target androgen binding site of the AR. While administration of these drugs can initially suppress the tumor growth, long-term therapy becomes progressively less effective. Among others, factors that make the AR less sensitive to conventional anti-androgens include 'gain-of-function' mutations (such as T877A and F876L) of the androgen binding site that convert AR antagonists to act as agonists, further contributing to cancer progression. Hence, there is an urgent need to develop new types of therapeutics that exhibit entirely different modes of AR inhibition and circumvent the drug resistance problem. One promising strategy for combating the mutation-driven resistance is to target alternative sites on the AR (such as BF3 regulatory pocket) and directly disrupt critical receptor-coactivator interactions that are essential for AR activity.

In this study, two computational strategies were applied to develop such AR BF3 inhibitors by taking the advantage of available resources of protein structures and known ligands. a) structure-based alone and b) a combination of structure- and ligand-based approaches. Both strategies yielded promising candidate molecules and novel chemo-types. Importantly, the developed AR BF3 inhibitors described in Chapters 3, 4 and 5 are structurally distinct from the commercially available anti-androgens of Enzalutamide series and nonspecific BF3 binders originally reported by Fletterick and colleagues. [63]

The developed drug candidates demonstrate direct selective interaction with the BF3 site, confirming this AR region as druggable. Our group has previously utilized the power of virtual screening combined with experimental evaluations to discover a number of small-molecules that

effectively target the BF3 site of the AR.[139] Taking the advantage of crystallographic structure of one of the identified inhibitors, **4035** (pdb id-2YLO),[139] we further developed a series of derivatives that belong to benzimidazole class. A simplified yet more active derivative **9002** was synthesized, experimentally evaluated and crystallographically resolved inside the AR BF3 target cavity (pdb id-4HLW). Another line of confirmation on this class of BF3 inhibitors was provided by a study conducted by Jehle *et al*.[66] They reported that compound **9114 (**a BF3 inhibitor discussed in chapter 3) successfully disrupted activating interactions between AR and Bag1L protein. The only limitation of the benzoimidazole series was that they did not achieve a level of potency that could be translated into significant *in vivo* implications. Hence, we initiated another round of systematic lead optimization that involved similarity searching and high-powered computational modeling as described in chapter 4. The identified compound **13163** belongs to indole chemical series that is confirmed to be effective in both *in vitro* assays and *in vivo* xenograft models.

During the lead optimization process, we observed that docking-based scoring functions failed to yield a statistically significant correlation between predicted docking scores and experimental activities of the identified chemicals. Hence, a part of this thesis work also focused on developing a method, termed *proteo-chemo-metric* approach to overcome the limitation of computational scoring functions. Chapter 5 described the development of new set of structure-activity descriptors that captured the sum of atomic distances between specific atomic type of active site residues and the docked ligands. These descriptors are straightforward to interpret and have never before been used in scoring functions. The QSAR models built on them demonstrated excellent predictability compared to commonly used scoring functions, highlighting the need for better structure-activity metrics. The developed models enabled us to prioritize further med.chem derivatives prior to their chemical synthesis. This approach was effective in terms of reducing time and costs involved during lead optimization process.

Further optimization of **13163** resulted in the development of indole-quinoline series of AR BF3 binders. One such compound - **13566** has been identified as the most potent BF3 inhibitor developed so far with its *in vitro* activity comparable to that of Enzalutamide. Importantly, the observations from molecular dynamic studies performed on AR-**13163**, AR-**13566** and AR-Bag1L motif complexes allowed us to elaborate the mechanism behind the possible displacement of GARRPR motif from the binding pocket in the presence of a BF3

inhibitor. These compounds perfectly occupy the binding regions of hexapeptide motif and disrupt key interactions between the Bag1L protein and BF3 residues. Florescence polarization and TR FRET assays confirmed that our BF3 inhibitors displace Bag1L protein effectively.

Drug resistance remains a fundamental cause of therapeutic failure in PCa therapy. This creates a challenge for researchers in the field of PCa drug discovery. It has been estimated that approximately 30% of CRPC patients harbor AR mutations [197] with most of them located within the ligand binding domain. Specific mutations that result in antagonist-to-agonist switch have been found recurrently at positions 877 and 741 in the LBD in samples from patients with metastatic CRPC, who have been treated with Hydroxyflutamide and Bicalutamide, respectively. Recently, a novel, third point mutation in the LBD (F876L) has been reported which developed upon treatment with Enzalutamide in PCa cell lines and xenograft models.[97, 102, 154] That particular F876L mutation is known to hamper the efficacy of Enzalutamide and its potent derivative ARN509.[198] Overall, these observations emphasize the urgency of developing the next generation of AR antagonists that provide therapeutic benefits in clinical applications.

Since BF3–directed compounds exhibit a novel mode of AR inhibition, we anticipated that they should address the issue of Enzalutamide resistance. Results from the cell viability assays indicated that **13566** exhibited strong anti-proliferative effect against LNCaP and Enzalutamide-resistant MR49F cell lines containing T877A and F876L forms of AR, respectively. No significant effect was detected on the growth of PC3 PCa cells which lack the AR entirely, which provides another indication of selective AR-directed effect of the developed BF3 inhibitors. The effectiveness of **13566** was further confirmed when it was shown to reduce the endogenous expression levels of PSA in Enzalutamide-resistant cells. More importantly, **13566** demonstrated significant inhibition of PCa tumors compared to Enzalutamide in LNCaP *in vivo* model. Recently, Lallous *et al* at Vancouver Prostate Centre investigated the effect of **13566** on 23 AR mutations identified from 62 CRPC patients. It has been observed that all these mutants demonstrate signs of resistance to at least one of commercial/experimental anti-androgens. The study highlighted that **13566** has been very effective against all 23 AR mutants and could and significantly inactivate the AR signaling axis in all the tested conditions (data not shown).[199] Overall, these results substantiate that targeting BF3 pocket is a viable option for PCa therapy. Especially when Enzalutamide resistance occurs, **treatment with AR BF3 based drugs**

could potentially provide another line of treatment and thereby enhance patient survival and slow disease progression.

It should be noted, that although BF3-directed drugs offer excellent therapeutic benefits, their applicability might be restricted in certain scenarios. These chemicals may be ineffective against AR splice variants that lack the ligand binding domain. However, it has been reported that AR variants tend to hetero-dimerize with full-length AR.[200] Furthermore, Watson *et al* demonstrated that some variants promote castration resistance by acting through full-length AR.[201] In a similar study, Xu *et al* reported that AR-V7 and AR^{v567es} not only homo-dimerize and hetero-dimerize with each other but also hetero-dimerize with full-length AR in an androgen-independent manner.[202] The fact that full-length AR is required for the function of variants validates the rationale behind developing BF3 inhibitors as an alternative therapeutic strategy for treating PCa.

6.2 Future Directions

In this study, we presented a cheminformatics scoring function that was trained and evaluated on the basis of a dataset of AR BF3 inhibitors. The developed method generally outperformed traditional docking-based scoring functions. In its current state, the scoring function presented is useful for predicting and prioritizing the derivatives prior to their chemical synthesis. However, a couple of improvements are feasible for future exploration.

6.2.1. First, the descriptors were developed based on the ligand poses generated by molecular docking method, which do not account for dynamic variations of the protein-ligand complex. Hence, introducing receptor flexibility in a docking protocol is a promising strategy to account for conformational changes induced by ligand binding. *Multiple receptor conformers* docking approach should help us to obtain more accurate geometry of the ligands docked into the BF3 pocket. Next, continuous algorithms (such as partial least squares) should be used along with machine-learning techniques to generate QSAR models. The advantage of linear QSAR methods is that they predict a range of activity rather than simply predicting whether a compound is active or inactive.

6.2.2. A key component in drug discovery is the process of achieving the optimum combination of potency and stability of a drug-like candidate. Although **13566** confirmed to be as potent as Enzalutamide, it still needs to be optimized for enhanced microsomal stability ($t_{1/2}$).

Increased $t_{1/2}$ should enable lower and less frequent dosing, thus promoting improved patient compliance. Therefore, further lead optimization led to the development of **13789** (a derivative of **13566**) with improved pharmacokinetic profile ($t_{1/2=}$ 269 mins). Further profiling of this compound is currently underway. This further opens up an opportunity to develop a clinical agent.

6.2.3. It has been proven that **13566** binds to BF3 site and inhibits AR activity with a mechanism that is entirely different from conventional anti-androgens. However, this mechanism needs to be studied in detail. It should be interesting to investigate if BF3 binders block nuclear translocation of AR and/or explore the panel of co-regulatory proteins that are critical for the AR transcriptional activity. A major clinical impact of the BF3 inhibitors would be if they may be used in combination with current anti-androgens to possibly avoid or delay progression to castration resistance. Therefore, it is worth testing the synergistic effect of **13566** and Enzalutamide in both *in vitro* assays and *in vivo* models.

6.2.4. Finally, the ultimate goal of developing any drug candidate is to make it orally available. Therefore studies directed towards developing oral formulations of these compounds could be an important part of future investigations.

References

1. Damber, J.-E.; Aus, G. Prostate cancer. *Lancet* **2008,** 371, 1710-21.
2. seer.cancer.gov.
3. Wilson, K. M.; Giovannucci, E. L.; Mucci, L. A. Lifestyle and dietary factors in the prevention of lethal prostate cancer. *Asian Journal of Andrology* **2012,** 14, 365-74.
4. Tilley, W. D.; Limtio, S. S.; Horsfall, D. J.; Aspinall, J. O.; Marshall, V. R.; Skinner, J. M. Detection of discrete androgen receptor epitopes in prostate-cancer by immunostaining - measurement by color video image-analysis. *Cancer Research* **1994,** 54, 4096-102.
5. Taplin, M. E.; Rajeshkumar, B.; Halabi, S.; Werner, C. P.; Woda, B. A.; Picus, J.; Stadler, W.; Hayes, D. F.; Kantoff, P. W.; Vogelzang, N. J.; Small, E. J. Androgen receptor mutations in androgen-in dependent prostate cancer: Cancer and Leukemia Group B Study 9663. *Journal of Clinical Oncology* **2003,** 21, 2673-8.
6. Vanderkwast, T. H.; Schalken, J.; Dewinter, J. A. R.; Vanvroonhoven, C. C. J.; Mulder, E.; Boersma, W.; Trapman, J. Androgen receptors in endocrine-therapy-resistant human prostate-cancer. *International Journal of Cancer* **1991,** 48, 189-93.
7. Mohler, J. L.; Chen, Y. Q.; Hamil, K.; Hall, S. H.; Cidlowski, J. A.; Wilson, E. M.; French, F. S.; Sar, M. Androgen and glucocorticoid receptors in the stroma and epithelium of prostatic hyperplasia and carcinoma. *Clinical Cancer Research* **1996,** 2, 889-95.
8. Hobisch, A.; Culig, Z.; Radmayr, C.; Bartsch, G.; Klocker, H.; Hittmair, A. Androgen receptor status of lymph node metastases from prostate cancer. *Prostate* **1996,** 28, 129-35.
9. Cheng, L.; Montironi, R.; Bostwick, D. G.; Lopez-Beltran, A.; Berney, D. M. Staging of prostate cancer. *Histopathology* **2012,** 60, 87-117.
10. Hayes, J. H.; Barry, M. J. Screening for Prostate Cancer With the Prostate-Specific Antigen Test A Review of Current Evidence. *Jama-Journal of the American Medical Association* **2014,** 311, 1143-9.
11. Pierorazio, P. M.; Walsh, P. C.; Partin, A. W.; Epstein, J. I. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *Bju International* **2013,** 111, 753-60.
12. Rukstalis, D. B. Treatment options after failure of radiation therapy-a review. *Reviews in urology* **2002,** 4 Suppl 2, S12-7.
13. Perlmutter, M. A.; Lepor, H. Androgen deprivation therapy in the treatment of advanced prostate cancer. *Reviews in urology* **2007,** 9 Suppl 1, S3-8.
14. Lassi, K.; Dawson, N. A. Emerging therapies in castrate-resistant prostate cancer. *Current Opinion in Oncology* **2009,** 21, 260-5.
15. Hotte, S. J.; Saad, F. Current management of castrate-resistant prostate cancer. *Current Oncology* **2010,** 17, S72-S9.
16. Saad, F.; Hotte, S. J.; Cuog; Cua. Guidelines for the management of castrate-resistant prostate cancer. *Cuaj-Canadian Urological Association Journal* **2010,** 4, 380-4.
17. Ye, L.; Kynaston, H. G.; Jiang, W. G. Bone metastasis in prostate cancer: Molecular and cellular mechanisms (Review). *International Journal of Molecular Medicine* **2007,** 20, 103-11.
18. Sturge, J.; Caley, M. P.; Waxman, J. Bone metastasis in prostate cancer: emerging therapeutic strategies. *Nature Reviews Clinical Oncology* **2011,** 8, 357-68.
19. Dutt, S. S.; Gao, A. C. Molecular mechanisms of castration-resistant prostate cancer progression. *Future Oncology* **2009,** 5, 1403-13.

20. Amaral, T. M. S.; Macedo, D.; Fernandes, I.; Costa, L. Castration-resistant prostate cancer: mechanisms, targets, and treatment. *Prostate cancer* **2012,** 2012, 327253-.
21. Gao, W. Q.; Bohl, C. E.; Dalton, J. T. Chemistry and structural biology of androgen receptor. *Chemical Reviews* **2005,** 105, 3352-70.
22. Johansen, K. L. Testosterone metabolism and replacement therapy in patients with end-stage renal disease. *Seminars in Dialysis* **2004,** 17, 202-8.
23. Kochakian, C. D. Comparison of protein anabolic property of various androgens in the castrated rat. *American Journal of Physiology* **1950,** 160, 53-61.
24. Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schutz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. The nuclear receptor superfamily - the 2nd decade. *Cell* **1995,** 83, 835-9.
25. Tsai, M. J.; Omalley, B. W. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annual Review of Biochemistry* **1994,** 63, 451-86.
26. Keller, E. T.; Ershler, W. B.; Chang, C. The androgen receptor: a mediator of diverse responses. *Frontiers in bioscience : a journal and virtual library* **1996,** 1, d59-71.
27. Yong, E. L.; Lim, J.; Qi, W.; Ong, V.; Mifsud, A. Molecular basis of androgen receptor diseases. *Annals of Medicine* **2000,** 32, 15-22.
28. Rosner, W.; Hryb, D. J.; Khan, M. S.; Nakhla, A. M.; Romas, N. A. Sex hormone-binding globulin - anatomy and physiology of a new regulatory system. *Journal of Steroid Biochemistry and Molecular Biology* **1991,** 40, 813-20.
29. Baker, M. E. Albumin, steroid hormones, and the origin of vertebrates. *American Zoologist* **2001,** 41, 1384-.
30. Georget, V.; Terouanne, B.; Nicolas, J. C.; Sultan, C. Mechanism of antiandrogen action: Key role of Hsp90 in conformational change and transcriptional activity of the androgen receptor. *Biochemistry* **2002,** 41, 11824-31.
31. Cutress, M. L.; Whitaker, H. C.; Mills, I. G.; Stewart, M.; Neal, D. E. Structural basis for the nuclear import of the human androgen receptor. *Journal of Cell Science* **2008,** 121, 957-68.
32. Wong, C. I.; Zhou, Z. X.; Sar, M.; Wilson, E. M. Steroid requirement for androgen receptor dimerization and DNA-binding - modulation by intramolecular interactions between the NH2-terminal and steroid-binding domains. *Journal of Biological Chemistry* **1993,** 268, 19004-12.
33. Zhou, X. E.; Suino-Powell, K. M.; Li, J.; He, Y.; MacKeigan, J. P.; Melcher, K.; Yong, E.-L.; Xu, H. E. Identification of SRC3/AIB1 as a Preferred Coactivator for Hormone-activated Androgen Receptor. *Journal of Biological Chemistry* **2010,** 285, 9161-71.
34. van Royen, M. E.; van Cappellen, W. A.; de Vos, C.; Houtsmuller, A. B.; Trapman, J. Stepwise androgen receptor dimerization. *Journal of Cell Science* **2012,** 125, 1970-9.
35. Bain, D. L.; Heneghan, A. F.; Connaghan-Jones, K. D.; Miura, M. T. Nuclear receptor structure: Implications for function. In *Annual Review of Physiology*, 2007; Vol. 69, pp 201-20.
36. Lavery, D. N.; McEwan, I. J. Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. *Biochemical Journal* **2005,** 391, 449-64.
37. McEwan, I. J. Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain. *Endocrine-Related Cancer* **2004,** 11, 281-93.
38. Lavery, D. N.; McEwan, I. J. The human androgen receptor AF1 transactivation domain: interactions with transcription factor IIF and molten-globule-like structural characteristics. *Biochemical Society Transactions* **2006,** 34, 1054-7.

39. Reid, J.; Betney, R.; Watt, K.; McEwan, I. J. The androgen receptor transactivation domain: the interplay between protein conformation and protein-protein interactions. *Biochemical Society Transactions* **2003,** 31, 1042-6.
40. Jenster, G.; Vanderkorput, H.; Trapman, J.; Brinkmann, A. O. Identification of 2 transcription activation units in the n-terminal domain of the human androgen receptor. *Journal of Biological Chemistry* **1995,** 270, 7341-6.
41. He, B.; Kemppainen, J. A.; Wilson, E. M. FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *Journal of Biological Chemistry* **2000,** 275, 22986-94.
42. He, B.; Wilson, E. M. The NH2-terminal and carboxyl-terminal interaction in the human androgen receptor. *Molecular Genetics and Metabolism* **2002,** 75, 293-8.
43. Laspada, A. R.; Wilson, E. M.; Lubahn, D. B.; Harding, A. E.; Fischbeck, K. H. Androgen receptor gene-mutations in x-linked spinal and bulbar muscular-atrophy. *Nature* **1991,** 352, 77-9.
44. Davies, P.; Watt, K.; Kelly, S. M.; Clark, C.; Price, N. C.; McEwan, I. J. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *Journal of Molecular Endocrinology* **2008,** 41, 301-14.
45. Betney, R.; McEwan, I. J. Role of conserved hydrophobic amino acids in androgen receptor AF-1 function. *Journal of Molecular Endocrinology* **2003,** 31, 427-39.
46. McEwan, I. J.; Gustafsson, J. A. Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** 94, 8485-90.
47. Fronsdal, K.; Engedal, N.; Slagsvold, T.; Saatcioglu, F. CREB binding protein is a coactivator for the androgen receptor and mediates cross-talk with AP-1. *Journal of Biological Chemistry* **1998,** 273, 31853-9.
48. Aarnisalo, P.; Palvimo, J. J.; Janne, O. A. CREB-binding protein in androgen receptor-mediated signaling. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** 95, 2122-7.
49. Dotzlaw, H.; Moehren, U.; Mink, S.; Cato, A. C. B.; Lluhi, J. A. I.; Baniahmad, A. The amino terminus of the human AR is target for corepressor action and antihormone agonism. *Molecular Endocrinology* **2002,** 16, 661-73.
50. Umesono, K.; Evans, R. M. Determinants of target gene specificity for steroid thyroid-hormone receptors. *Cell* **1989,** 57, 1139-46.
51. Claessens, F.; Alen, P.; Devos, A.; Peeters, B.; Verhoeven, G.; Rombauts, W. The androgen-specific probasin response element 2 interacts differentially with androgen and glucocorticoid receptors. *Journal of Biological Chemistry* **1996,** 271, 19013-6.
52. Denayer, S.; Helsen, C.; Thorrez, L.; Haelens, A.; Claessens, F. The Rules of DNA Recognition by the Androgen Receptor. *Molecular Endocrinology* **2010,** 24, 898-913.
53. Shaffer, P. L.; Jivan, A.; Dollins, D. E.; Claessens, F.; Gewirth, D. T. Structural basis of androgen receptor binding to selective androgen response elements. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** 101, 4758-63.
54. Gaughan, L.; Stockley, J.; Wang, N.; McCracken, S. R. C.; Treumann, A.; Armstrong, K.; Shaheen, F.; Watt, K.; McEwan, I. J.; Wang, C.; Pestell, R. G.; Robson, C. N. Regulation of the androgen receptor by SET9-mediated methylation. *Nucleic Acids Research* **2011,** 39, 1266-79.

55. Gaughan, L.; Logan, I. R.; Neal, D. E.; Robson, C. N. Regulation of androgen receptor and histone deacetylase 1 by Mdm2-mediated ubiquitylation. *Nucleic Acids Research* **2005,** 33, 13-26.
56. Zhou, Z. X.; Sar, M.; Simental, J. A.; Lane, M. V.; Wilson, E. M. Ligand-dependent bipartite nuclear targeting signal in the human androgen receptor - requirement for the DNA-binding domain and modulation by NH2-terminal and carboxyl-terminal sequences. *Journal of Biological Chemistry* **1994,** 269, 13115-23.
57. Carsonjurica, M. A.; Schrader, W. T.; Omalley, B. W. Steroid-receptor family-Strucutre and functions. *Endocrine Reviews* **1990,** 11, 201-20.
58. Pereira de Jesus-Tran, K.; Cote, P.-L.; Cantin, L.; Blanchet, J.; Labrie, F.; Breton, R. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein science : a publication of the Protein Society* **2006,** 15, 987-99.
59. He, B.; Gampe, R. T.; Kole, A. J.; Hnat, A. T.; Stanley, T. B.; An, G.; Stewart, E. L.; Kalman, R. I.; Minges, J. T.; Wilson, E. M. Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Molecular Cell* **2004,** 16, 425-38.
60. Hur, E.; Pfaff, S. J.; Payne, E. S.; Gron, H.; Buehrer, B. M.; Fletterick, R. J. Recognition and accommodation at the androgen receptor coactivator binding interface. *Plos Biology* **2004,** 2, 1303-12.
61. He, B.; Kemppainen, J. A.; Voegel, J. J.; Gronemeyer, H.; Wilson, E. M. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH2-terminal domain. *Journal of Biological Chemistry* **1999,** 274, 37219-25.
62. van Royen, M. E.; Cunha, S. M.; Brink, M. C.; Mattern, K. A.; Nigg, A. L.; Dubbink, H. J.; Verschure, P. J.; Trapman, J.; Houtsmuller, A. B. Compartmentalization of androgen receptor protein-protein interactions in living cells. *Journal of Cell Biology* **2007,** 177, 63-72.
63. Estebanez-Perpina, E.; Arnold, A. A.; Nguyen, P.; Rodrigues, E. D.; Mar, E.; Bateman, R.; Pallai, P.; Shokat, K. M.; Baxter, J. D.; Guy, R. K.; Webb, P.; Fletterick, R. J. A surface on the androgen receptor that allosterically regulates coactivator binding. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, 16074-9.
64. Munuganti, R. S. N.; Leblanc, E.; Axerio-Cilies, P.; Labriere, C.; Frewin, K.; Singh, K.; et al. Targeting the Binding Function 3 (BF3) Site of the Androgen Receptor Through Virtual Screening. 2. Development of 2-((2-phenoxyethyl) thio)-1H-benzimidazole Derivatives. *J Med Chem* **2013,** 56, 1136-48.
65. De Leon, J. T.; Iwai, A.; Feau, C.; Garcia, Y.; Balsiger, H. A.; Storer, C. L.; Suro, R. M.; Garza, K. M.; Lee, S.; Kim, Y. S.; Chen, Y.; Ning, Y. M.; Riggs, D. L.; Fletterick, R. J.; Guy, R. K.; Trepel, J. B.; Neckers, L. M.; Cox, M. B. Targeting the regulation of androgen receptor signaling by the heat shock protein 90 cochaperone FKBP52 in prostate cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** 108, 11878-83.
66. Jehle, K.; Cato, L.; Neeb, A.; Muhle-Goll, C.; Jung, N.; Smith, E.; Buzon, V.; Carbó, L.; Estébanez-Perpiñá, E.; Schmitz, K.; Fruk, L.; Luy, B.; Chen, Y.; Cox, M.; Bräse, S.; Brown, M.; Cato, A. Coregulator control of androgen receptor action by a novel nuclear receptor-binding motif. *J Biol Chem* **2014,** 289, 8839-51.

67. Gottlieb, B.; Beitel, L. K.; Nadarajah, A.; Paliouras, M.; Trifiro, M. The Androgen Receptor Gene Mutations Database: 2012 Update. *Human Mutation* **2012,** 33, 887-94.
68. Fang, S.; Liao, S. Antagonistic action of anti-androgens on formation of a specific dihydrotestosterone-receptor protein complex in rat ventral prostate. *Molecular Pharmacology* **1969,** 5, 428-&.
69. Gao, W.; Kim, J.; Dalton, J. T. Pharmacokinetics and pharmacodynamics of nonsteroidal androgen receptor ligands. *Pharmaceutical Research* **2006,** 23, 1641-58.
70. Cockshott, I. D. Bicalutamide - Clinical pharmacokinetics and metabolism. *Clinical Pharmacokinetics* **2004,** 43, 855-78.
71. Tran, C.; Ouk, S.; Clegg, N. J.; Chen, Y.; Watson, P. A.; Arora, V.; Wongvipat, J.; Smith-Jones, P. M.; Yoo, D.; Kwon, A.; Wasielewska, T.; Welsbie, D.; Chen, C. D.; Higano, C. S.; Beer, T. M.; Hung, D. T.; Scher, H. I.; Jung, M. E.; Sawyers, C. L. Development of a Second-Generation Antiandrogen for Treatment of Advanced Prostate Cancer. *Science* **2009,** 324, 787-90.
72. Clegg, N. J.; Wongvipat, J.; Joseph, J. D.; Tran, C.; Ouk, S.; Dilhas, A.; Chen, Y.; Grillot, K.; Bischoff, E. D.; Cal, L.; Aparicio, A.; Dorow, S.; Arora, V.; Shao, G.; Qian, J.; Zhao, H.; Yang, G. B.; Cao, C. Y.; Sensintaffar, J.; Wasielewska, T.; Herbert, M. R.; Bonnefous, C.; Darimont, B.; Scher, H. I.; Smith-Jones, P.; Klang, M.; Smith, N. D.; De Stanchina, E.; Wu, N.; Ouerfelli, O.; Rix, P. J.; Heyman, R. A.; Jung, M. E.; Sawyers, C. L.; Hager, J. H. ARN-509: A Novel Antiandrogen for Prostate Cancer Treatment. *Cancer Research* **2012,** 72, 1494-503.
73. Potter, G. A.; Barrie, S. E.; Jarman, M.; Rowlands, M. G. Novel steroidal inhibitors of human cytochrome p450(17-alpha) (17-alpha-hydroxylase-C-17,C-20-lyase) - potential agents for the treatment of prostatic-cancer. *Journal of Medicinal Chemistry* **1995,** 38, 2463-71.
74. Barrie, S. E.; Haynes, B. P.; Potter, G. A.; Chan, F. C. Y.; Goddard, P. M.; Dowsett, M.; Jarman, M. Biochemistry and pharmacokinetics of potent non-steroidal cytochrome P450(17 alpha) inhibitors. *Journal of Steroid Biochemistry and Molecular Biology* **1997,** 60, 347-51.
75. De Bono, J. S.; Logothetis, C. J.; Molina, A.; Fizazi, K.; North, S.; Chu, L.; Chi, K. N.; Jones, R. J.; Goodman, O. B., Jr.; Saad, F.; Staffurth, J. N.; Mainwaring, P.; Harland, S.; Flaig, T. W.; Hutson, T. E.; Cheng, T.; Patterson, H.; Hainsworth, J. D.; Ryan, C. J.; Sternberg, C. N.; Ellard, S. L.; Flechon, A.; Saleh, M.; Scholz, M.; Efstathiou, E.; Zivi, A.; Bianchini, D.; Loriot, Y.; Chieffo, N.; Thian, K.; Haqq, C. M.; Scher, H. I.; Investigators, C.-A.-. Abiraterone and Increased Survival in Metastatic Prostate Cancer. *New England Journal of Medicine* **2011,** 364, 1995-2005.
76. Handratta, V. D.; Vasaitis, T. S.; Njar, V. C. O.; Gediya, L. K.; Kataria, R.; Chopra, P.; Newman, D.; Farquhar, R.; Guo, Z. Y.; Qiu, Y.; Brodie, A. M. H. 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. *Journal of Medicinal Chemistry* **2005,** 48, 2972-84.
77. Yu, Z.; Cai, C.; Gao, S.; Simon, N. I.; Shen, H. C.; Balk, S. P. Galeterone Prevents Androgen Receptor Binding to Chromatin and Enhances Degradation of Mutant Androgen Receptor. *Clinical Cancer Research* **2014,** 20, 4075-85.
78. Fizazi, K.; Massard, C.; Bono, P.; Jones, R.; Kataja, V.; James, N.; Garcia, J. A.; Protheroe, A.; Tammela, T. L.; Elliott, T.; Mattila, L.; Aspegren, J.; Vuorela, A.; Langmuir, P.; Mustonen, M.; Grp, A. S. Activity and safety of ODM-201 in patients with progressive metastatic castration-resistant prostate cancer (ARADES): an open-label phase 1 dose-escalation and randomised phase 2 dose expansion trial. *Lancet Oncology* **2014,** 15, 975-85.

79. Narayanan, R.; Mohler, M. L.; Bohl, C. E.; Miller, D. D.; Dalton, J. T. Selective androgen receptor modulators in preclinical and clinical development. *Nuclear receptor signaling* **2008,** 6, e010-e.
80. Chen, J. Y.; Kim, J.; Dalton, J. T. Discovery and therapeutic promise of selective androgen receptor modulators. *Molecular Interventions* **2005,** 5, 173-88.
81. Dalton, J. T.; Mukherjee, A.; Zhu, Z. X.; Kirkovsky, L.; Miller, D. D. Discovery of nonsteroidal androgens. *Biochemical and Biophysical Research Communications* **1998,** 244, 1-4.
82. He, Y. L.; Yin, D. H.; Perera, M.; Kirkovsky, L.; Stourman, N.; Li, W.; Dalton, J. T.; Miller, D. D. Novel nonsteroidal ligands with high binding affinity and potent functional activity for the androgen receptor. *European Journal of Medicinal Chemistry* **2002,** 37, 619-34.
83. Srinath, R.; Dobs, A. Enobosarm (GTx-024, S-22): a potential treatment for cachexia. *Future Oncology* **2014,** 10, 187-94.
84. Nickols, N. G.; Dervan, P. B. Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, 10418-23.
85. Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. Cyclic Pyrrole-Imidazole Polyamides Targeted to the Androgen Response Element. *Journal of the American Chemical Society* **2009,** 131, 7182-8.
86. Cherian, M. T.; Wilson, E. M.; Shapiro, D. J. A Competitive Inhibitor That Reduces Recruitment of Androgen Receptor to Androgen-responsive Genes. *Journal of Biological Chemistry* **2012,** 287, 23368-80.
87. Li, H.; Ban, F.; Dalal, K.; Leblanc, E.; Frewin, K.; Ma, D.; Adomat, H.; Rennie, P. S.; Cherkasov, A. Discovery of Small-Molecule Inhibitors Selectively Targeting the DNA-Binding Domain of the Human Androgen Receptor. *Journal of Medicinal Chemistry* **2014,** 57, 6458-67.
88. Dalal, K.; Roshan-Moniri, M.; Sharma, A.; Li, H.; Ban, F.; Hessein, M.; Hsing, M.; Singh, K.; LeBlanc, E.; Dehm, S.; Guns, E. S. T.; Cherkasov, A.; Rennie, P. S. Selectively Targeting the DNA-binding Domain of the Androgen Receptor as a Prospective Therapy for Prostate Cancer. *Journal of Biological Chemistry* **2014,** 289, 26417-29.
89. Sadar, M. D. Small Molecule Inhibitors Targeting the "Achilles' Heel" of Androgen Receptor Activity. *Cancer Research* **2011,** 71, 1208-13.
90. Sadar, M. D.; Williams, D. E.; Mawji, N. R.; Patrick, B. O.; Wikanta, T.; Chasanah, E.; Irianto, H. E.; Van Soest, R.; Andersen, R. J. Sintokamides A to E, Chlorinated Peptides from the Sponge Dysidea sp that Inhibit Transactivation of the N-Terminus of the Androgen Receptor in Prostate Cancer Cells. *Organic Letters* **2008,** 10, 4947-50.
91. Brand, L. J.; Olson, M. E.; Ravindranathan, P.; Guo, H.; Kempema, A. M.; Andrews, T. E.; Chen, X.; Raj, G. V.; Harki, D. A.; Dehm, S. M. EPI-001 is a selective peroxisome proliferator-activated receptor-gamma modulator with inhibitory effects on androgen receptor expression and activity in prostate cancer. *Oncotarget* **2015,** 6, 3811-24.
92. Quayle, S. N.; Mawji, N. R.; Wang, J.; Sadar, M. D. Androgen receptor decoy molecules block the growth of prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** 104, 1331-6.
93. Chen, C. D.; Welsbie, D. S.; Tran, C.; Baek, S. H.; Chen, R.; Vessella, R.; Rosenfeld, M. G.; Sawyers, C. L. Molecular determinants of resistance to antiandrogen therapy. *Nature Medicine* **2004,** 10, 33-9.

94. Visakorpi, T.; Hyytinen, E.; Koivisto, P.; Tanner, M.; Keinanen, R.; Palmberg, C.; Palotie, A.; Tammela, T.; Isola, J.; Kallioniemi, O. P. In-vivo amplification of the androgen receptor gene and progression of human prostate-cancer. *Nature Genetics* **1995,** 9, 401-6.
95. Koivisto, P.; Kononen, J.; Palmberg, C.; Tammela, T.; Hyytinen, E.; Isola, J.; Trapman, J.; Cleutjens, K.; Noordzij, A.; Visakorpi, T.; Kallioniemi, O. P. Androgen receptor gene amplification: A possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Research* **1997,** 57, 314-9.
96. Matias, P. M.; Donner, P.; Coelho, R.; Thomaz, M.; Peixoto, C.; Macedo, S.; Otto, N.; Joschko, S.; Scholz, P.; Wegg, A.; Basler, S.; Schafer, M.; Egner, U.; Carrondo, M. A. Structural evidence for ligand specificity in the binding domain of the human androgen receptor - Implications for pathogenic gene mutations. *Journal of Biological Chemistry* **2000,** 275, 26164-71.
97. Balbas, M. D.; Evans, M. J.; Hosfield, D. J.; Wongvipat, J.; Arora, V. K.; Watson, P. A.; Chen, Y.; Greene, G. L.; Shen, Y.; Sawyers, C. L. Overcoming mutation-based resistance to antiandrogens with rational drug design. *eLife* **2013,** 2, e00499-e.
98. Hara, T.; Miyazaki, J.; Araki, H.; Yamaoka, M.; Kanzaki, N.; Kusaka, M.; Miyamoto, M. Novel mutations of androgen receptor: A possible mechanism of bicalutamide withdrawal syndrome. *Cancer Research* **2003,** 63, 149-53.
99. Bohl, C. E.; Gao, W. Q.; Miller, D. D.; Bell, C. E.; Dalton, J. T. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** 102, 6201-6.
100. Bohl, C. E.; Wu, Z.; Miller, D. D.; Bell, C. E.; Dalton, J. T. Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *Journal of Biological Chemistry* **2007,** 282, 13648-55.
101. Sun, C.; Shi, Y.; Xu, L. L.; Nageswararao, C.; Davis, L. D.; Segawa, T.; Dobi, A.; McLeod, D. G.; Srivastava, S. Androgen receptor mutation (T877A) promotes prostate cancer cell growth and cell survival. *Oncogene* **2006,** 25, 3905-13.
102. Korpal, M.; Korn, J. M.; Gao, X.; Rakiec, D. P.; Ruddy, D. A.; Doshi, S.; Yuan, J.; Kovats, S. G.; Kim, S.; Cooke, V. G.; Monahan, J. E.; Stegmeier, F.; Roberts, T. M.; Sellers, W. R.; Zhou, W.; Zhu, P. An F876L Mutation in Androgen Receptor Confers Genetic and Phenotypic Resistance to MDV3100 (Enzalutamide). *Cancer Discovery* **2013,** 3, 1030-43.
103. Dehm, S. M.; Schmidt, L. J.; Heemers, H. V.; Vessella, R. L.; Tindall, D. J. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Research* **2008,** 68, 5469-77.
104. Hu, R.; Dunn, T. A.; Wei, S.; Isharwal, S.; Veltri, R. W.; Humphreys, E.; Han, M.; Partin, A. W.; Vessella, R. L.; Isaacs, W. B.; Bova, G. S.; Luo, J. Ligand-Independent Androgen Receptor Variants Derived from Splicing of Cryptic Exons Signify Hormone-Refractory Prostate Cancer. *Cancer Research* **2009,** 69, 16-22.
105. Ahrens-Fath, I.; Politz, O.; Geserick, C.; Haendler, B. Androgen receptor function is modulated by the tissue-specific AR45 variant. *Febs Journal* **2005,** 272, 74-84.
106. Guo, Z.; Yang, X.; Sun, F.; Jiang, R.; Linn, D. E.; Chen, H.; Chen, H.; Kong, X.; Melamed, J.; Tepper, C. G.; Kung, H.-J.; Brodie, A. M. H.; Edwards, J.; Qiu, Y. A Novel Androgen Receptor Splice Variant Is Up-regulated during Prostate Cancer Progression and Promotes Androgen Depletion-Resistant Growth. *Cancer Research* **2009,** 69, 2305-13.

107. Hornberg, E.; Ylitalo, E. B.; Crnalic, S.; Antti, H.; Stattin, P.; Widmark, A.; Bergh, A.; Wikstrom, P. Expression of Androgen Receptor Splice Variants in Prostate Cancer Bone Metastases is Associated with Castration-Resistance and Short Survival. *Plos One* **2011,** 6.
108. Li, Y.; Chan, S. C.; Brand, L. J.; Hwang, T. H.; Silverstein, K. A. T.; Dehm, S. M. Androgen Receptor Splice Variants Mediate Enzalutamide Resistance in Castration-Resistant Prostate Cancer Cell Lines. *Cancer Research* **2013,** 73, 483-9.
109. Gregory, C. W.; Fei, X. Y.; Ponguta, L. A.; He, B.; Bill, H. M.; French, F. S.; Wilson, E. M. Epidermal growth factor increases coactivation of the androgen receptor in recurrent prostate cancer. *Journal of Biological Chemistry* **2004,** 279, 7119-30.
110. Krueckl, S. L.; Sikes, R. A.; Edlund, N. M.; Bell, R. H.; Hurtado-Coll, A.; Fazli, L.; Gleave, M. E.; Cox, M. E. Increased insulin-like growth factor I receptor expression and signaling are components of androgen-independent progression in a lineage-derived prostate cancer progression model. *Cancer Research* **2004,** 64, 8620-9.
111. Hobisch, A.; Eder, I. E.; Putz, T.; Horninger, W.; Bartsch, G.; Klocker, H.; Culig, Z. Interleukin-6 regulates prostate specific protein expression in prostate carcinoma cells by activation of the androgen receptor. *Cancer Research* **1998,** 58, 4640-5.
112. Asim, M.; Siddiqui, I. A.; Hafeez, B. B.; Baniahmad, A.; Mukhtar, H. Src kinase potentiates androgen receptor transactivation function and invasion of androgen-independent prostate cancer C4-2 cells. *Oncogene* **2008,** 27, 3596-604.
113. Carver, B. S.; Chapinski, C.; Wongvipat, J.; Hieronymus, H.; Chen, Y.; Chandarlapaty, S.; Arora, V. K.; Le, C.; Koutcher, J.; Scher, H.; Scardino, P. T.; Rosen, N.; Sawyers, C. L. Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer. *Cancer Cell* **2011,** 19, 575-86.
114. Edlind, M. P.; Hsieh, A. C. PI3K-AKT-mTOR signaling in prostate cancer progression and androgen deprivation therapy resistance. *Asian Journal of Andrology* **2014,** 16, 378-86.
115. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* **2003,** 22, 151-85.
116. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *British Journal of Pharmacology* **2011,** 162, 1239-49.
117. Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology* **2007,** 25, 71-5.
118. Kurosawa, G.; Akahori, Y.; Morita, M.; Sumitomo, M.; Sato, N.; Muramatsu, C.; Eguchi, K.; Matsuda, K.; Takasaki, A.; Tanaka, M.; Iba, Y.; Hamada-Tsutsumi, S.; Ukaie, Y.; Shiraishi, M.; Suzuki, K.; Kurosawa, M.; Fujiyama, S.; Takahashi, N.; Kato, R.; Mizoguchi, Y.; Shamoto, M.; Tsuda, H.; Sugiurak, M.; Hattori, Y.; Miyakawa, S.; Shiroki, R.; Hoshinaga, K.; Hayashi, N.; Sugioka, A.; Kurosawa, Y. Comprehensive screening for antigens overexpressed on carcinomas via isolation of human mAbs that may be therapeutic. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, 7287-92.
119. Fox, S.; Farr-Jones, S.; Sopchak, L.; Boggs, A.; Nicely, H. W.; Khoury, R.; Biros, M. High-throughput screening: Update on practices and success. *Journal of Biomolecular Screening* **2006,** 11, 864-9.
120. Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery* **2004,** 3, 711-5.

121. Polson, A. G.; Fuji, R. N. The successes and limitations of preclinical studies in predicting the pharmacodynamics and safety of cell-surface-targeted biological agents in patients. *British Journal of Pharmacology* **2012,** 166, 1600-2.
122. Collier, R. Rapidly rising clinical trial costs worry researchers. *Canadian Medical Association Journal* **2009,** 180, 277-8.
123. Oprea, T. I. Current trends in lead discovery: Are we looking for the appropriate properties? *Molecular Diversity* **2000,** 5, 199-208.
124. Cherkasov, A.; Ban, F.; Santos-Filho, O.; Thorsteinson, N.; Fallahi, M.; Hammond, G. L. An updated steroid benchmark set and its application in the discovery of novel nanomolar ligands of sex hormone-binding globulin. *Journal of Medicinal Chemistry* **2008,** 51, 2047-56.
125. Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004,** 303, 1813-8.
126. Hou, T. J.; Xu, X. J. Recent development and application of virtual screening in drug discovery: An overview. *Current Pharmaceutical Design* **2004,** 10, 1011-33.
127. Talele, T. T.; Khedkar, S. A.; Rigby, A. C. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Current Topics in Medicinal Chemistry* **2010,** 10, 127-41.
128. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **2004,** 47, 1739-49.
129. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics & Modelling* **2007,** 26, 198-212.
130. Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor-sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology* **1995,** 245, 43-53.
131. Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *Bmc Biology* **2011,** 9.
132. Williams, C. Reverse fingerprinting, similarity searching by group fusion and fingerprint bit importance. *Molecular Diversity* **2006,** 10, 311-32.
133. MACCS structural keys. Symyx Software, San Ramon; 2002.
134. Barnard, J. M.; Downs, G. M. Chemical fragment generation and clustering software. *Journal of Chemical Information and Computer Sciences* **1997,** 37, 141-2.
135. Jaccard, P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles* **1901,** 37.
136. Todeschini R; V, C. Handbook of molecular descriptors. *Weinheim: Wiley-VCH* **2000,** 11.
137. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry* **2014,** 57, 4977-5010.
138. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* **2010,** 15, 444-50.
139. Lack, N. A.; Axerio-Cilies, P.; Tavassoli, P.; Han, F. Q.; Chan, K. H.; Feau, C.; LeBlanc, E.; Guns, E. T.; Guy, R. K.; Rennie, P. S.; Cherkasov, A. Targeting the Binding Function 3

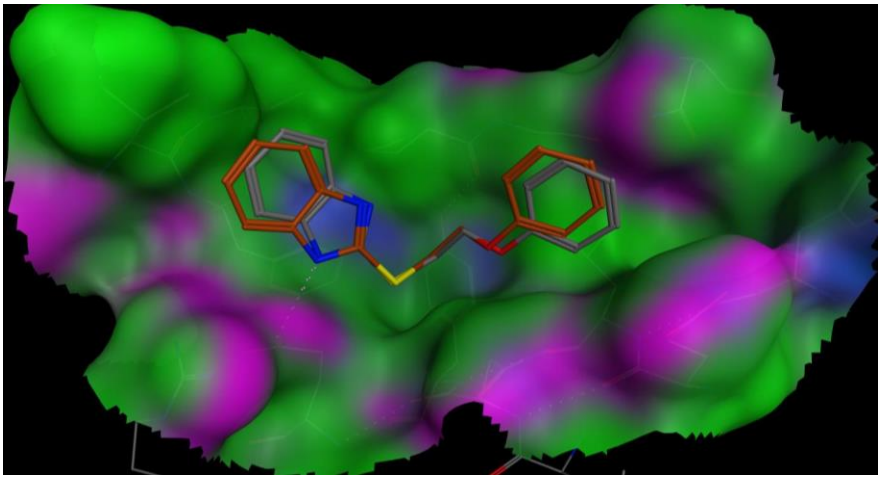
(BF3) Site of the Human Androgen Receptor through Virtual Screening. *Journal of Medicinal Chemistry* **2011,** 54, 8563-73.
140. Chemical Computing Group Inc. Molecular Operating Environment (MOE); Montreal, Quebec, Canada, 2008; . www. chemcomp.com.
141. Schrodinger. Maestro; Schrodinger: New York, 2008;. www. schrodinger.com.
142. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design* **1997,** 11, 425-45.
143. Instant JChem 5.10, ChemAxon. http://www.chemaxon.com
144. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling* **2012,** 52, 1757-68.
145. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997,** 101, 3005-14.
146. Munuganti, R.; Hassona, M.; Leblanc, E.; Frewin, K.; Ma, D.; Ban, F.; Hsing, M.; Andre, C.; Jonadass, J.; Zoubeidi, A.; Young, R.; Guns, E.; Rennie, P.; Cherkasov, A. Development of New Potent Anti-Androgens that Target Binding Function-3 (BF3) Site as an Alternative Therapeutic Option to Treat Castration Resistant Prostate Cancer. *Journal of Medicinal Chemistry* **2014,** Under Revision.
147. Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for 3-dimensional structure-directed quantitative structure-activity-relationships .1. partition-coefficients as a measure of hydrophobicity. *Journal of Computational Chemistry* **1986,** 7, 565-77.
148. Kramer, C.; Gedeck, P. Global Free Energy Scoring Functions Based on Distance-Dependent Atom-Type Pair Descriptors. *Journal of Chemical Information and Modeling* **2011,** 51, 707-20.
149. Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I. H. Data mining in bioinformatics using Weka. *Bioinformatics* **2004,** 20, 2479-81.
150. Roy, K. On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opinion on Drug Discovery* **2007,** 2, 1567-77.
151. Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences* **2003,** 43, 579-86.
152. Zou, K. H.; O'Malley, A. J.; Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **2007,** 115, 654-7.
153. Tavassoli, P.; Snoek, R.; Ray, M.; Rao, L. G.; Rennie, P. S. Rapid, non-destructive, cell-based screening assays for agents that modulate growth, death, and androgen receptor activation in prostate cancer cells. *Prostate* **2007,** 67, 416-26.
154. Kuruma, H.; Matsumoto, H.; Shiota, M.; Bishop, J.; Lamoureux, F.; Thomas, C.; Briere, D.; Los, G.; Gleave, M.; Fanjul, A.; Zoubeidi, A. A Novel Antiandrogen, Compound 30, Suppresses Castration-Resistant and MDV3100-Resistant Prostate Cancer Growth In Vitro and In Vivo. *Molecular Cancer Therapeutics* **2013,** 12, 567-76.
155. Tirat, A.; Freuler, F.; Stettler, T.; Mayr, L. M.; Leder, L. Evaluation of two novel tag-based labelling technologies for site-specific modification of proteins. *International Journal of Biological Macromolecules* **2006,** 39, 66-76.

156. Balk, S. P.; Ko, Y. J.; Bubley, G. J. Biology of prostate-specific antigen. *Journal of Clinical Oncology* **2003,** 21, 383-91.
157. Horoszewicz, J. S.; Leong, S. S.; Chu, T. M.; Wajsman, Z. L.; Friedman, M.; Papsidero, L.; Kim, U.; Chai, L. S.; Kakati, S.; Arya, S. K.; Sandberg, A. A. The LNCaP cell line--a new model for studies on human prostatic carcinoma. *Progress in clinical and biological research* **1980,** 37, 115-32.
158. Kaighn, M. E.; Narayan, K. S.; Ohnuki, Y.; Lechner, J. F.; Jones, L. W. Establishment and characterization of a human prostatic-carcinoma cell-line (PC-3). *Investigative Urology* **1979,** 17, 16-23.
159. Arkin, M. R.; Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. *Nature Reviews Drug Discovery* **2004,** 3, 301-17.
160. Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of small molecules to an adaptive protein-protein interface. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** 100, 1603-8.
161. Szakacs, G.; Paterson, J. K.; Ludwig, J. A.; Booth-Genthe, C.; Gottesman, M. M. Targeting multidrug resistance in cancer. *Nature Reviews Drug Discovery* **2006,** 5, 219-34.
162. Seruga, B.; Ocana, A.; Tannock, I. F. Drug resistance in metastatic castration-resistant prostate cancer. *Nature Reviews Clinical Oncology* **2011,** 8, 12-23.
163. Pavlidis, P.; Noble, W. S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **2003,** 19, 295-6.
164. Laudet, V. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *Journal of Molecular Endocrinology* **1997,** 19, 207-26.
165. Buzon, V.; Carbo, L. R.; Estruch, S. B.; Fletterick, R. J.; Estebanez-Perpina, E. A conserved surface on the ligand binding domain of nuclear receptors for allosteric control. *Molecular and Cellular Endocrinology* **2012,** 348, 394-402.
166. Miyake, H.; Nelson, C.; Rennie, P. S.; Gleave, M. E. Acquisition of chemoresistant phenotype by overexpression of the antiapoptotic gene testosterone-repressed prostate message-2 in prostate cancer xenograft models. *Cancer Research* **2000,** 60, 2547-54.
167. Sato, N.; Gleave, M. E.; Bruchovsky, N.; Rennie, P. S.; Goldenberg, S. L.; Lange, P. H.; Sullivan, L. D. Intermittent androgen suppression delays progression to androgen-independent regulation of prostate-specific: Antigen gene in the LNCaP prostate tumour model. *Journal of Steroid Biochemistry and Molecular Biology* **1996,** 58, 139-46.
168. Cheng, H.; Snoek, R.; Ghaidi, F.; Cox, N. E.; Rennie, P. S. Short hairpin RNA knockdown of the androgen receptor attenuates ligand-independent activation and delays tumor progression. *Cancer Research* **2006,** 66, 10613-20.
169. Snoek, R.; Cheng, H.; Margiotti, K.; Wafa, L. A.; Wong, C. A.; Wong, E. C.; Fazli, L.; Nelson, C. C.; Gleave, M. E.; Rennie, P. S. In vivo Knockdown of the Androgen Receptor Results in Growth Inhibition and Regression of Well-Established, Castration-Resistant Prostate Tumors. *Clinical Cancer Research* **2009,** 15, 39-47.
170. Zhang, K. X.; Moussavi, M.; Kim, C.; Chow, E.; Chen, I. S.; Fazli, L.; Jia, W.; Rennie, P. S. Lentiviruses with trastuzumab bound to their envelopes can target and kill prostate cancer cells. *Cancer Gene Therapy* **2009,** 16, 820-31.
171. Leeson, P.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery* **2007,** 6, 881-90.

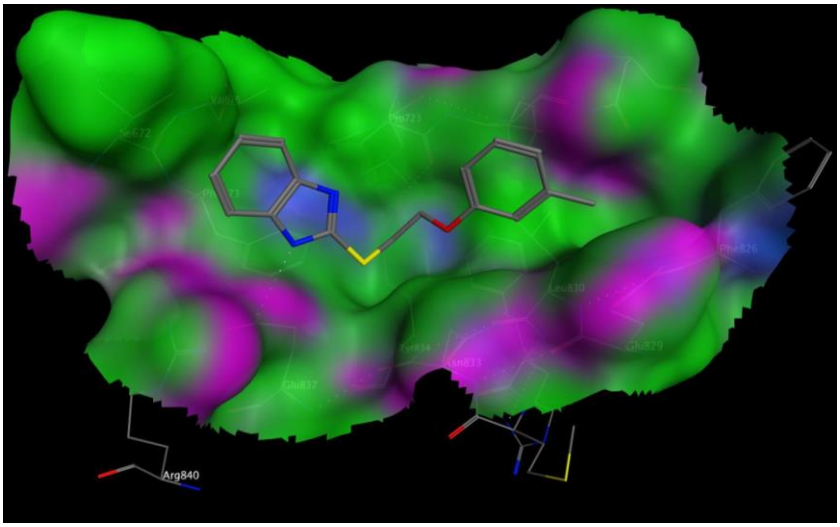
172. Gleave, M. E.; Hsieh, J. T.; Wu, H. C.; Voneschenbach, A. C.; Chung, L. W. K. Serum prostate specific antigen levels in mice bearing human prostate LNCaP tumors are determined by tumor volume and endocrine and growth factors. . *Cancer Research* **1992,** 52, 1598-605.
173. Blundell, T. L. Structure-based drug design. *Nature* **1996,** 384, 23-6.
174. Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992,** 257, 1078-82.
175. Seifert, M. H. J.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Current Opinion in Drug Discovery & Development* **2007,** 10, 298-307.
176. Bohm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *Journal of Computer-Aided Molecular Design* **1998,** 12, 309-23.
177. Joseph-McCarthy, D.; Baber, J. C.; Feyfant, E.; Thompson, D. C.; Humblet, C. Lead optimization via high-throughput molecular docking. *Current Opinion in Drug Discovery & Development* **2007,** 10, 264-74.
178. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery* **2004,** 3, 935-49.
179. Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* **2006,** 49, 5912-31.
180. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *Journal of Medicinal Chemistry* **2006,** 49, 5851-5.
181. Coupez, B.; Lewis, R. A. Docking and scoring - Theoretically easy, practically impossible? *Current Medicinal Chemistry* **2006,** 13, 2995-3003.
182. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein-ligand interactions. *Journal of Medicinal Chemistry* **2004,** 47, 3032-47.
183. Englebienne, P.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 5. Force-Field-Based Prediction of Binding Affinities of Ligands to Proteins. *Journal of Chemical Information and Modeling* **2009,** 49, 2564-71.
184. Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* **2000,** 295, 337-56.
185. Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry* **2005,** 48, 6296-303.
186. Muegge, I. PMF scoring revisited. *Journal of Medicinal Chemistry* **2006,** 49, 5895-902.
187. Mooij, W. T. M.; Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins-Structure Function and Bioinformatics* **2005,** 61, 272-87.
188. Pham, T. A.; Jain, A. N. Customizing scoring functions for docking. *Journal of Computer-Aided Molecular Design* **2008,** 22, 269-86.
189. Lapinsh, M.; Prusis, P.; Gutcaits, A.; Lundstedt, T.; Wikberg, J. E. S. Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochimica Et Biophysica Acta-General Subjects* **2001,** 1525, 180-90.
190. Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A. A new approach to the theoretical estimation of inductive constants. *Journal of Physical Organic Chemistry* **1998,** 11, 437-47.

191. Cherkasov, A. R.; Galkin, V.; Cherkasov, R. "Inductive" electronegativity scale. *Journal of Molecular Structure-Theochem* **1999,** 489, 43-6.
192. Cherkasov, A. Inductive electronegativity scale. Iterative calculation of inductive partial charges. *Journal of Chemical Information and Computer Sciences* **2003,** 43, 2039-47.
193. Cherkasov, A.; Jankovic, B. Application of 'inductive' QSAR descriptors for quantification of antibacterial activity of cationic polypeptides. *Molecules* **2004,** 9, 1034-52.
194. Jenssen, H.; Lejon, T.; Hilpert, K.; Fjell, C. D.; Cherkasov, A.; Hancock, R. E. W. Evaluating different descriptors for model design of antimicrobial peptides with enhanced activity toward P-aeruginosa. *Chemical Biology & Drug Design* **2007,** 70, 134-42.
195. Smith, R. D.; Dunbar, J. B.; Ung, P. M. U.; Esposito, E. X.; Yang, C. Y.; Wang, S. M.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Combined Evaluation Across All Submitted Scoring Functions. *Journal of Chemical Information and Modeling* **2011,** 51, 2115-31.
196. Huggins, C.; Stevens, R. E.; Hodges, C. V. Studies on prostate cancer II The effects of castration on advanced carcinoma of the prostate gland. *Archives of Surgery* **1941,** 43, 209-23.
197. Waltering, K. K.; Urbanucci, A.; Visakorpi, T. Androgen receptor (AR) aberrations in castration-resistant prostate cancer. *Molecular and Cellular Endocrinology* **2012,** 360, 38-43.
198. Joseph, J. D.; Lu, N.; Qian, J.; Sensintaffar, J.; Shao, G.; Brigham, D.; Moon, M.; Maneval, E. C.; Chen, I.; Darimont, B.; Hager, J. H. A clinically relevant androgen receptor mutation confers resistance to second-generation antiandrogens enzalutamide and ARN-509. *Cancer Discovery* **2013,** 3, 1020-9.
199. Lallous, N.; Volik, S.; Awrey, S.; LeBlanc, E.; Tse, R.; Murillo, J.; Singh, K.; Azad, A.; Wyatt, A.; LeBihan, S.; Chi, K.; Gleave, M.; Rennie, P.; Collins, C.; Cherkasov, A. Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. *Genome Biology* **2015,** Accepted Manuscript.
200. Qi, Y.; Cao, B.; Zhang, G.; Xu, D.; Guo, Z.; Xiong, Z.; Plymate, S.; Sartor, O.; Zhang, H.; Dong, Y. Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy. *Cancer Research* **2014,** 74.
201. Watson, P. A.; Chen, Y. F.; Balbas, M. D.; Wongvipat, J.; Socci, N. D.; Viale, A.; Kim, K.; Sawyers, C. L. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** 107, 16759-65.
202. Xu, D.; Zhan, Y.; Qi, Y. F.; Cao, B.; Bai, S. S.; Xu, W.; Gambhir, S. S.; Lee, P.; Sartor, O.; Flemington, E. K.; Zhang, H. T.; Hu, C. D.; Dong, Y. Androgen Receptor Splice Variants Dimerize to Transactivate Target Genes. *Cancer Research* **2015,** 75, 3663-71.

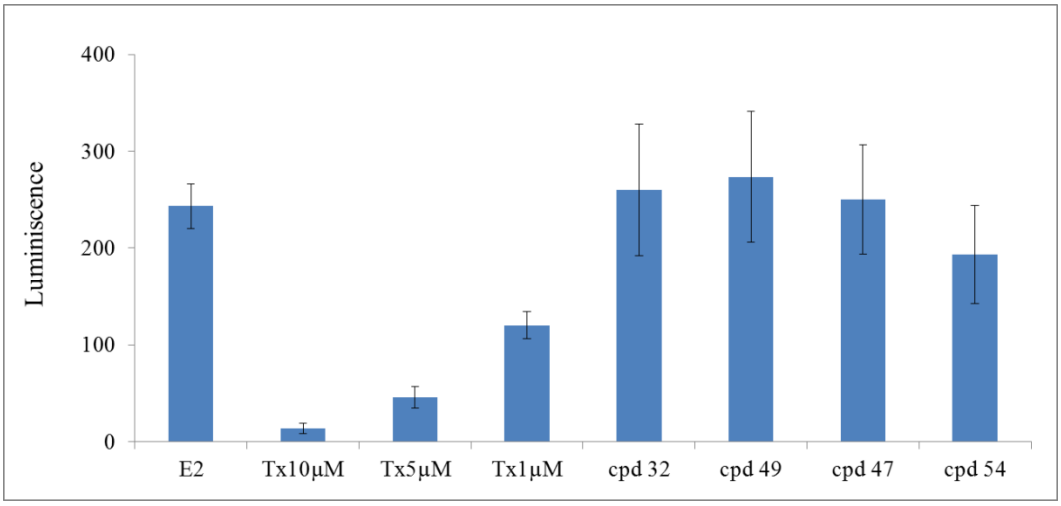
Appendix



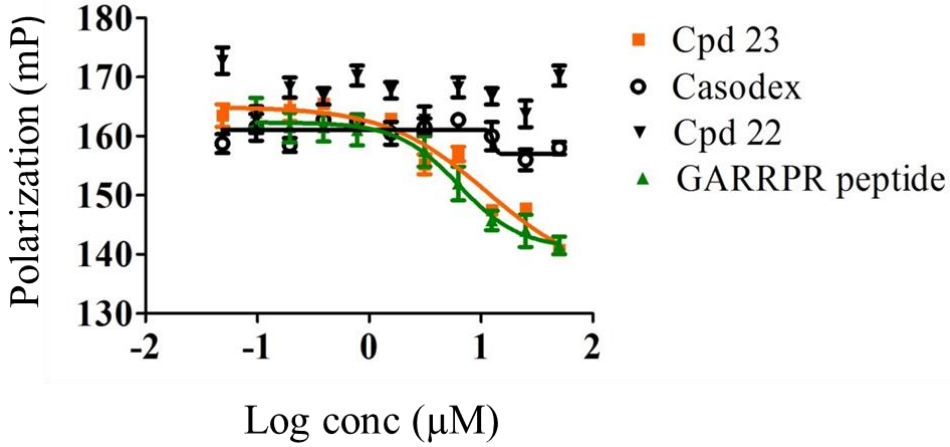
Supplementary figure 3.1. Predicted docking pose (grey molecule) of compound **9002** versus experimentally (brown molecule) identified BF3-bound configuration. White dotted lines indicate hydrogen bond with Glu837.



Supplementary figure 3.2. Predicted docking pose (grey molecule) of compound **9006**. White dotted lines indicate hydrogen bond with Glu837.

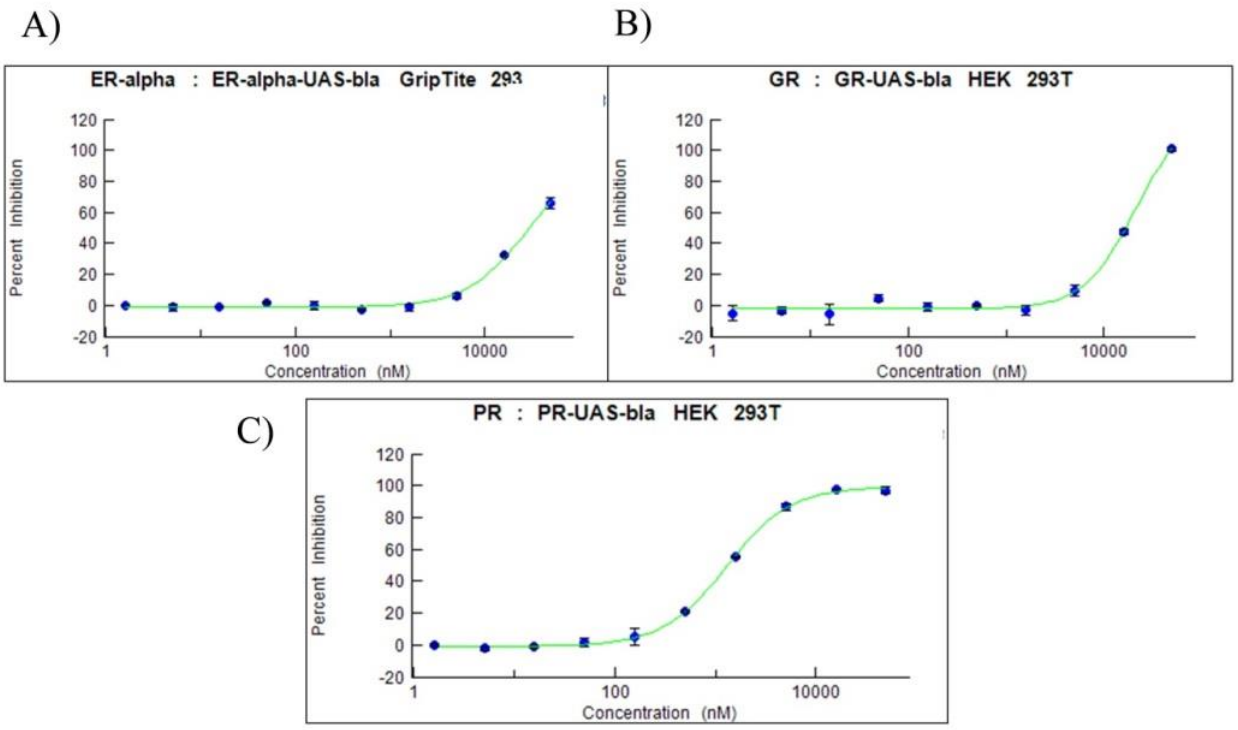


Supplementary figure 3.3. ERα mediated luciferase transcriptional assay showing the specificity of AR BF3 compounds. The compounds (10 μM) were unable to inhibit estrogen response element mediated transcriptional activity compared to Tamoxifen (Tx). For each compound the luminescence signal for 17βEstradiol stimulated MCF-7 cells is shown. The error bars represent standard deviation for n=6 values.

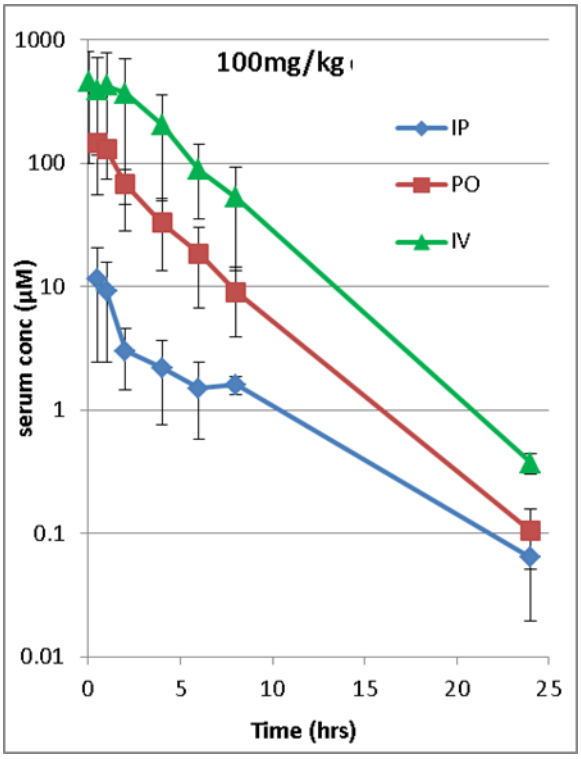


Supplementary figure 4.1. Fluorescence polarization experiment showing competition between FITC-labeled Bag-1L (61-80) peptide (12.5nM) and serially diluted Casodex, **13163**, **13221**, unlabeled core GARRPR hexapaptide (100-0.05μM) for binding to mutated AR-LBD-T877A (2

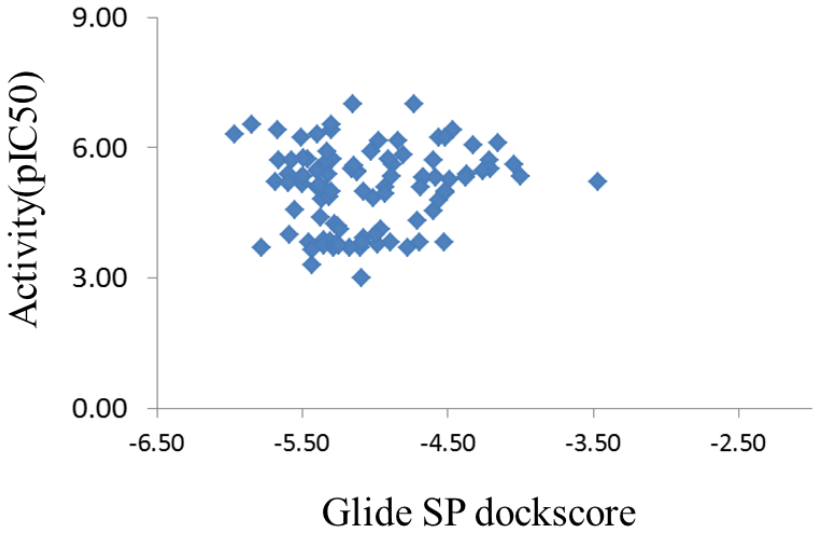
μM). The competition experiments were performed at conditions for maximum polarization of FITC-labeled peptide and AR-LBD.



Supplementary figure 4.2. The effect of 13163 on A) Estrogen receptor alpha B) Glucocorticoid receptor C) Progesterone receptor, as measured by Life Technologies SELECT screen. The experiment was performed using HEK cells expressing the Gal4 DNA binding domain fused to the ligand binding domain of AR, ER, PR or GR according to Life Technologies protocol.



Supplementary figure 4.3. 13163 Inhibits AR-dependent Growth of Xenograft Tumors *In Vivo*. Pharmacokinetic profiles of compound 13163 following intravenous (IV), oral (PO) or intraperitoneal (IP) administration at 100mg/kg (n=3 each). IV and oral dosing are observed to be the best routes of delivery. For IV, PO and IP respectively, Cmax was approximately 400, 150 and 12µM and half-life 2.3, 2.3 and 3.4 hours.



Supplementary figure 5.1. The correlation curve plotted between the experimental activity (inhibition of AR activity as measured by AR eGFP assay) and dock score produced by Glide SP program for 106 BF3 inhibitors. The graph shows that there is no correlation observed.

Supplemental table 4.1. The list of compounds selected through virtual screening process as described in chapter 4. The scores predicted by different *in silico* protocols (Glide score, eHiTS score, RMSD, LigX pKi) for each compound and their cumulative vote are reported below.

ID	Glide SP-Score	eHiTS-Score	RMSD	LigX	dock_pKi	SP vote	eHiTS vote	RMSD vote	Lig X vote	pki vote	Final Vote
13259	-5.301	-5.805	0.612	-28.894	4.890	1	1	1	0	1	4
ZINC00297315	-5.364	-5.479	0.460	-30.843	5.246	1	0	1	1	1	4
13163	-5.850	-5.800	0.624	-32.878	4.952	1	0	1	1	1	4
13256	-5.306	-5.600	0.512	-31.733	4.900	1	0	1	1	1	4
ZINC00019965	-5.517	-5.783	0.606	-32.516	2.621	1	1	1	1	0	4
13309	-5.554	-6.040	1.353	-30.793	4.850	1	1	0	1	1	4
13127	-5.510	-5.493	0.642	-32.141	4.840	1	0	1	1	1	4
13221	-5.371	-5.769	1.155	-35.253	4.821	1	1	0	1	1	4
ZINC00079157	5.640	-6.316	0.698	-30.061	3.706	1	1	1	1	0	4
13310	-5.300	-5.841	1.051	-31.064	4.798	1	1	0	1	1	4
13299	-5.405	-6.105	1.476	-30.205	4.747	1	1	0	1	1	4
13167	-5.660	-5.871	0.530	-30.165	4.702	1	0	1	1	1	4
ZINC00273820	-5.731	-6.150	0.516	-29.302	5.090	1	1	1	0	1	4
13164	-5.975	-5.820	0.801	-31.196	4.850	1	1	0	1	1	4
13320	-5.267	-5.893	1.602	-32.107	4.697	1	1	0	1	1	4
ZINC17206741	-5.145	-5.766	1.668	-32.081	5.763	1	1	0	1	1	4
13225	-5.460	-5.790	0.390	-33.676	4.653	1	0	1	1	1	4
13257	-5.600	-5.977	1.593	-30.434	4.625	1	1	0	1	1	4
ZINC59513820	-5.391	-6.228	0.663	-31.554	4.203	1	1	1	1	0	4
13166	-5.490	-5.835	0.391	-31.603	4.604	1	0	1	1	1	4
13226	-5.156	-6.447	1.251	-31.233	4.533	1	1	0	1	1	4
13258	-5.500	-5.411	0.500	-30.987	4.477	1	0	1	1	1	4
ZINC05741193	-4.705	-5.799	0.642	-28.645	5.684	0	1	1	0	1	3
ZINC13125776	-5.320	-5.312	0.420	-25.954	5.646	1	0	1	0	1	3
13255	-5.407	-5.839	1.350	-29.970	4.889	1	1	0	0	1	3
ZINC00270881	-5.284	-5.597	0.622	-20.676	4.868	1	0	1	0	1	3
ZINC59518358	-5.180	-5.127	0.580	-26.198	4.867	1	0	1	0	1	3
13303	-5.492	-5.750	1.639	-30.124	4.796	1	0	0	1	1	3
13304	-5.387	-5.807	1.276	-29.994	4.700	1	1	0	0	1	3
13300	-5.330	-5.800	1.647	-30.369	4.640	1	0	0	1	1	3
ZINC21002281	-5.024	-5.383	0.651	-30.925	4.607	0	0	1	1	1	3
ZINC00287779	-5.209	-5.043	0.509	-27.005	4.593	1	0	1	0	1	3
13345	-5.390	-5.046	0.617	-31.368	4.497	1	0	1	1	0	3
13254	-5.330	-5.780	0.653	-31.531	4.485	1	0	1	1	0	3
ZINC59518222	-4.697	-5.832	0.527	-32.993	4.127	0	1	1	1	0	3
13321	-5.269	-5.989	0.783	-30.027	4.580	1	1	0	0	1	3
ZINC59513854	-4.871	-5.132	0.710	-28.983	5.437	0	0	1	0	1	2
ZINC00047188	-4.802	-5.289	0.700	-22.418	5.408	0	0	1	0	1	2
ZINC04756486	-5.206	-5.625	1.453	-24.439	4.950	1	0	0	0	1	2
ZINC03196421	-4.462	-5.741	1.145	-27.238	4.895	0	1	0	0	1	2
ZINC83388991	-3.681	-5.186	0.689	-23.937	4.883	0	0	1	0	1	2
ZINC00296941	-4.504	-5.785	1.884	-26.656	4.671	0	1	0	0	1	2
ZINC40570824	-5.347	-5.360	1.952	-21.548	4.664	1	0	0	0	1	2
ZINC72137657	-4.679	-5.157	0.627	-25.803	4.644	0	0	1	0	1	2
ZINC16525325	-5.275	-5.300	0.759	-28.758	4.629	1	0	0	0	1	2
ZINC02749872	-4.920	-5.947	1.400	-24.393	4.553	0	1	0	0	1	2
ZINC59248175	-5.285	-4.989	0.975	-26.864	4.539	1	0	0	0	1	2
ZINC00046098	-4.957	-5.279	1.988	-31.618	4.526	0	0	0	1	1	2
ZINC05619237	-5.101	-4.938	0.908	-31.198	4.363	1	0	0	1	0	2
ZINC75260628	-4.896	-6.178	0.460	-26.292	4.187	0	1	1	0	0	2
ZINC40570823	-4.468	-5.642	0.552	-30.937	4.181	0	0	1	1	0	2
ZINC72177086	-4.614	-6.293	1.318	-31.127	4.115	0	1	0	1	0	2
ZINC00587972	-4.674	-6.280	1.865	-31.550	3.987	0	1	0	1	0	2
ZINC00298052	-5.077	-5.950	1.918	-32.554	3.902	0	1	0	1	0	2

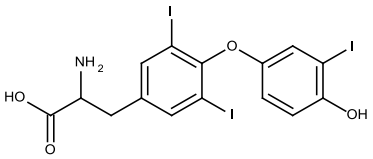
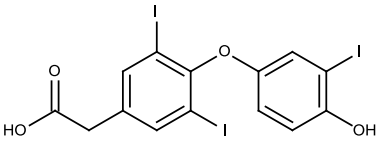
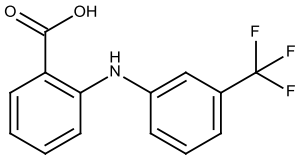
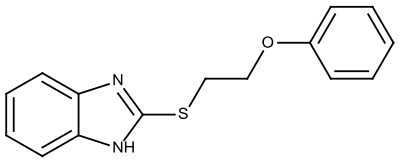
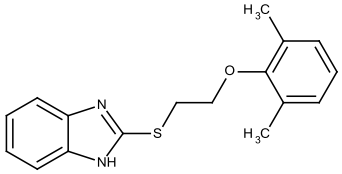
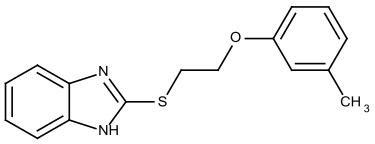
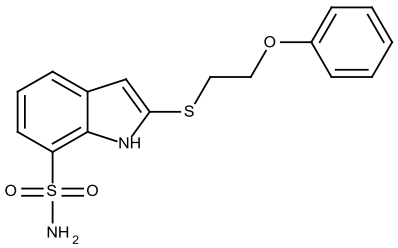
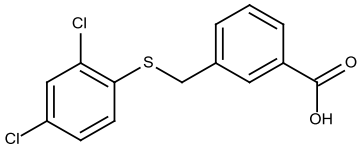
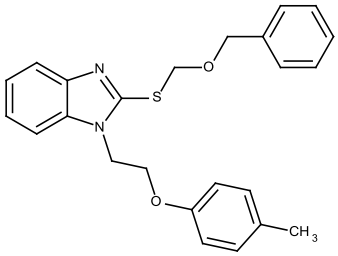
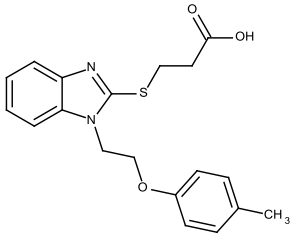
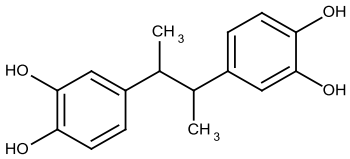
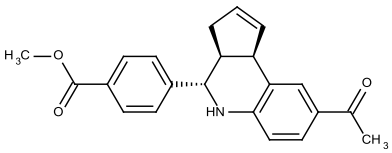
ID	Glide SP-Score	eHiTS-Score	RMSD	LigX	dock_pKi	SP vote	eHiTS vote	RMSD vote	Lig X vote	pki vote	Final Vote
ZINC03240772	-4.974	-6.327	1.789	-30.951	3.898	0	1	0	1	0	2
ZINC21002283	-5.271	-5.439	1.638	-31.551	3.851	1	0	0	1	0	2
ZINC00046422	-4.866	-5.814	0.770	-30.185	3.836	0	1	0	1	0	2
ZINC40570771	-5.374	-4.975	0.652	-21.125	3.830	1	0	1	0	0	2
ZINC18191571	-4.681	-6.117	2.650	-30.820	3.793	0	1	0	1	0	2
ZINC38805545	-5.303	-4.943	1.650	-31.765	3.752	1	0	0	1	0	2
ZINC59247885	-5.318	-4.996	1.500	-31.116	3.715	1	0	0	1	0	2
ZINC00242590	-4.836	-5.015	0.708	-31.739	3.662	0	0	1	1	0	2
ZINC00062723	-4.274	-5.921	0.703	-29.215	3.630	0	1	1	0	0	2
ZINC00584458	-4.429	-6.192	0.597	-28.567	3.627	0	1	1	0	0	2
ZINC00065287	-4.657	-5.755	0.951	-30.944	3.585	0	1	0	1	0	2
ZINC13144651	-5.132	-4.954	0.667	-29.615	3.509	1	0	1	0	0	2
ZINC00064459	-4.650	-5.577	0.715	-32.242	3.376	0	0	1	1	0	2
ZINC04713988	-4.153	-6.402	0.835	-31.219	3.039	0	1	0	1	0	2
ZINC04707741	-4.999	-5.074	0.600	-32.140	2.835	0	0	1	1	0	2
ZINC00273916	-3.858	-5.870	1.762	-30.599	2.598	0	1	0	1	0	2
ZINC04694547	-5.009	-5.000	1.098	-29.235	5.457	0	0	0	0	1	1
ZINC40452777	-4.210	-5.215	0.748	-28.017	5.342	0	0	0	0	1	1
ZINC40570831	-4.283	-5.692	1.416	-29.857	5.253	0	0	0	0	1	1
ZINC59249148	-4.759	-4.991	1.221	-26.188	5.215	0	0	0	0	1	1
ZINC04906320	-4.933	-5.392	1.100	-24.284	5.153	0	0	0	0	1	1
ZINC59386206	-4.318	-5.045	1.535	-26.691	5.013	0	0	0	0	1	1
ZINC05669183	-5.081	-5.323	1.218	-29.064	4.929	0	0	0	0	1	1
ZINC83388990	-3.695	-4.940	1.502	-27.455	4.858	0	0	0	0	1	1
ZINC59518550	-4.734	-5.153	0.884	-26.387	4.858	0	0	0	0	1	1
ZINC00065809	-4.777	-5.151	0.999	-27.786	4.848	0	0	0	0	1	1
ZINC18137618	-4.435	-5.157	1.289	-24.913	4.726	0	0	0	0	1	1
ZINC73733927	-4.442	-5.081	1.805	-24.903	4.702	0	0	0	0	1	1
ZINC00271359	-4.899	-5.281	1.357	-26.576	4.681	0	0	0	0	1	1
ZINC05536636	-5.097	-5.473	1.319	-26.002	4.636	0	0	0	0	1	1
ZINC59302900	-4.132	-4.995	1.870	-27.902	4.546	0	0	0	0	1	1
ZINC40570776	-4.263	-4.984	1.638	-28.205	4.543	0	0	0	0	1	1
ZINC00047177	-5.049	-5.438	1.900	-25.579	4.540	0	0	0	0	1	1
ZINC85442351	-4.997	-5.189	0.845	-23.675	4.523	0	0	0	0	1	1
ZINC13209067	-5.068	-5.001	0.932	-22.048	4.517	0	0	0	0	1	1
ZINC02541087	-5.039	-4.952	1.983	-24.926	4.511	0	0	0	0	1	1
ZINC03194754	-5.214	-5.353	1.112	-24.695	4.502	1	0	0	0	0	1
ZINC00502159	-5.225	-5.263	1.098	-28.415	4.497	1	0	0	0	0	1
ZINC59513845	-4.511	-5.276	0.673	-26.648	4.422	0	0	1	0	0	1
ZINC00051081	-5.360	-5.486	0.924	-27.846	4.409	1	0	0	0	0	1
ZINC85442353	-4.458	-5.191	0.459	-23.747	4.402	0	0	1	0	0	1
ZINC04728918	-4.575	-4.958	1.145	-30.638	4.373	0	0	0	1	0	1
ZINC00297515	-4.821	-6.354	1.804	-28.692	4.291	0	1	0	0	0	1
ZINC01274758	-4.838	-5.559	1.687	-33.489	4.284	0	0	0	1	0	1
ZINC59457042	-4.487	-5.155	0.990	-30.629	4.239	0	0	0	1	0	1
ZINC00136664	-5.279	-5.033	1.130	-28.092	4.228	1	0	0	0	0	1
ZINC71767002	-3.365	-4.962	0.414	-25.968	4.228	0	0	1	0	0	1
ZINC33653051	-4.474	-4.945	0.569	-26.746	4.212	0	0	1	0	0	1
ZINC40570697	-4.140	-5.389	0.673	-27.969	4.202	0	0	1	0	0	1
ZINC00439163	-4.808	-5.336	1.153	-36.827	4.184	0	0	0	1	0	1
ZINC00297619	-5.052	-5.835	1.598	-23.763	4.155	0	1	0	0	0	1
ZINC59487898	-4.456	-5.711	0.974	-30.435	4.154	0	0	0	1	0	1
ZINC04821853	-5.354	-5.069	1.710	-25.386	4.151	1	0	0	0	0	1
ZINC59519276	-3.971	-5.579	0.616	-20.158	4.129	0	0	1	0	0	1
ZINC00288470	-4.413	-5.264	0.712	-26.252	4.116	0	0	1	0	0	1
ZINC59457245	-5.162	-5.357	1.695	-25.419	4.094	1	0	0	0	0	1
ZINC59418781	-4.582	-5.766	1.120	-28.875	4.090	0	1	0	0	0	1
ZINC00576379	-4.886	-5.392	0.527	-29.927	4.000	0	0	1	0	0	1
ZINC00062467	-5.169	-5.422	1.095	-29.461	3.986	1	0	0	0	0	1
ZINC03250977	-4.959	-6.194	0.832	-15.536	3.978	0	1	0	0	0	1
ZINC00588219	-4.276	-5.456	0.803	-35.958	3.975	0	0	0	1	0	1

ID	Glide SP-Score	eHiTS-Score	RMSD	LigX	dock_pKi	SP vote	eHiTS vote	RMSD vote	Lig X vote	pki vote	Final Vote
ZINC00273957	-5.084	-5.332	0.696	-29.947	3.971	0	0	1	0	0	1
ZINC12346351	-4.164	-5.800	1.217	-29.860	3.971	0	1	0	0	0	1
ZINC03036427	-4.814	-5.033	0.635	-26.406	3.953	0	0	1	0	0	1
ZINC01274756	-5.256	-5.553	2.122	-28.574	3.943	1	0	0	0	0	1
ZINC39410466	-5.237	-5.495	1.641	-21.729	3.921	1	0	0	0	0	1
ZINC59678792	-5.252	-4.990	1.845	-27.021	3.915	1	0	0	0	0	1
ZINC00668188	-4.495	-5.762	1.628	-26.143	3.871	0	1	0	0	0	1
ZINC00588159	-5.345	-5.266	1.016	-26.936	3.861	1	0	0	0	0	1
ZINC00357299	-5.091	-5.522	1.650	-31.936	3.848	0	0	0	1	0	1
ZINC40570820	-5.110	-5.645	0.963	-27.116	3.783	1	0	0	0	0	1
ZINC03589754	-5.184	-5.173	1.781	-22.401	3.769	1	0	0	0	0	1
ZINC04713984	-4.746	-5.345	0.548	-29.324	3.768	0	0	1	0	0	1
ZINC09482118	-3.710	-5.006	0.738	-29.484	3.749	0	0	1	0	0	1
ZINC04821732	-4.561	-5.073	0.574	-25.115	3.739	0	0	1	0	0	1
ZINC03233313	-4.466	-6.269	1.007	-28.299	3.711	0	1	0	0	0	1
ZINC00068921	-5.039	-5.535	1.843	-32.167	3.694	0	0	0	1	0	1
ZINC13608197	-5.194	-5.116	1.901	-26.747	3.592	1	0	0	0	0	1
ZINC38174436	-4.597	-5.378	0.634	-27.493	3.581	0	0	1	0	0	1
ZINC00358027	-5.291	-5.678	0.943	-29.267	3.568	1	0	0	0	0	1
ZINC20508916	-4.268	-5.032	0.667	-27.110	3.566	0	0	1	0	0	1
ZINC39932995	-3.940	-5.412	1.808	-32.318	3.496	0	0	0	1	0	1
ZINC05517877	-4.990	-6.161	1.736	-26.666	3.484	0	1	0	0	0	1
ZINC03158604	-5.020	-5.766	1.812	-25.625	3.476	0	1	0	0	0	1
ZINC02131693	-5.149	-5.696	1.843	-28.589	3.474	1	0	0	0	0	1
ZINC59508156	-3.990	-5.988	1.121	-26.419	3.449	0	1	0	0	0	1
ZINC01086150	-4.205	-5.023	1.597	-30.501	3.440	0	0	0	1	0	1
ZINC59513935	-4.756	-5.851	1.394	-24.467	3.439	0	1	0	0	0	1
ZINC00243289	-4.474	-5.012	0.940	-31.061	3.438	0	0	0	1	0	1
ZINC17001876	-4.742	-4.957	0.736	-27.345	3.428	0	0	1	0	0	1
ZINC01052351	-4.662	-5.000	0.714	-28.706	3.370	0	0	1	0	0	1
ZINC00297828	-4.178	-6.260	0.986	-27.687	3.366	0	1	0	0	0	1
ZINC00576365	-4.639	-5.393	1.998	-33.355	3.312	0	0	0	1	0	1
ZINC18191561	-4.667	-5.110	0.741	-27.497	3.256	0	0	1	0	0	1
ZINC00282379	-4.648	-5.153	0.595	-27.058	3.254	0	0	1	0	0	1
ZINC03154475	-4.998	-5.038	0.401	-29.497	3.233	0	0	1	0	0	1
ZINC00079154	-4.861	-6.302	0.986	-24.577	3.221	0	1	0	0	0	1
ZINC05337327	-4.058	-4.966	0.620	-24.843	3.207	0	0	1	0	0	1
ZINC00497896	-4.660	-6.394	2.681	-24.592	3.205	0	1	0	0	0	1
ZINC01709001	-3.986	-5.440	0.843	-30.502	3.175	0	0	0	1	0	1
ZINC01044367	-3.795	-5.066	0.854	-30.093	3.165	0	0	0	1	0	1
ZINC04957827	-4.663	-6.337	2.231	-25.238	3.141	0	1	0	0	0	1
ZINC32620423	-4.122	-5.903	0.922	-23.749	3.115	0	1	0	0	0	1
ZINC00306551	-4.083	-5.080	0.622	-24.112	3.101	0	0	1	0	0	1
ZINC59418779	-4.797	-6.067	1.550	-29.156	3.073	0	1	0	0	0	1
ZINC00971280	-4.593	-5.149	1.358	-32.530	3.038	0	0	0	1	0	1
ZINC06269866	-5.171	-4.938	1.998	-25.974	3.020	1	0	0	0	0	1
ZINC00065652	-4.670	-6.015	1.035	-28.651	2.996	0	1	0	0	0	1
ZINC00943139	-4.551	-5.762	1.862	-23.401	2.948	0	1	0	0	0	1
ZINC24410425	-4.397	-5.113	0.451	-28.669	2.931	0	0	1	0	0	1
ZINC00571310	-4.539	-5.573	1.831	-32.026	2.898	0	0	0	1	0	1
ZINC00065058	-4.616	-5.012	1.102	-27.769	4.503	0	0	0	0	0	0
ZINC06269861	-3.947	-4.939	1.827	-26.924	4.503	0	0	0	0	0	0
ZINC00267862	-4.340	-5.373	0.778	-26.659	4.471	0	0	0	0	0	0
ZINC00065374	-3.441	-5.046	1.563	-27.653	4.462	0	0	0	0	0	0
ZINC21002279	-4.757	-5.401	2.404	-28.332	4.442	0	0	0	0	0	0
ZINC00271175	-4.168	-5.544	0.798	-25.983	4.441	0	0	0	0	0	0
ZINC39248931	-4.905	-5.391	1.209	-29.221	4.411	0	0	0	0	0	0
ZINC01274757	-4.150	-5.561	1.038	-27.501	4.410	0	0	0	0	0	0
ZINC73733923	-4.394	-5.086	0.801	-12.145	4.408	0	0	0	0	0	0
ZINC05977375	-4.358	-4.941	1.744	-28.254	4.382	0	0	0	0	0	0
ZINC04380137	-4.969	-5.338	1.917	-26.336	4.319	0	0	0	0	0	0

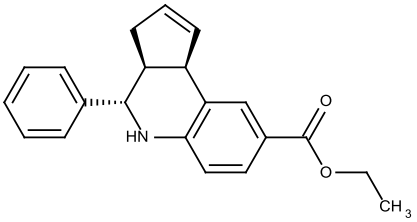
ID	Glide SP-Score	eHiTS-Score	RMSD	LigX	dock_pKi	SP vote	eHiTS vote	RMSD vote	Lig X vote	pki vote	Final Vote
ZINC05257525	-4.967	-5.117	0.759	-23.067	4.310	0	0	0	0	0	0
ZINC08616443	-5.024	-5.061	1.402	-23.224	4.306	0	0	0	0	0	0
ZINC06760081	-4.617	-5.198	1.311	-27.774	4.250	0	0	0	0	0	0
ZINC59513661	-4.098	-5.050	1.488	-24.437	4.235	0	0	0	0	0	0
ZINC18191566	-4.622	-5.723	1.694	-29.442	4.233	0	0	0	0	0	0
ZINC00061061	-4.858	-5.003	0.834	-25.078	4.213	0	0	0	0	0	0
ZINC01044365	-4.508	-5.065	1.658	-22.920	4.211	0	0	0	0	0	0
ZINC00054670	-4.890	-5.007	1.438	-24.277	4.197	0	0	0	0	0	0
ZINC72134786	-3.818	-5.068	0.928	-23.892	4.195	0	0	0	0	0	0
ZINC03158614	-4.110	-5.141	1.704	-26.912	4.176	0	0	0	0	0	0
ZINC89209310	-5.058	-5.118	1.502	-27.857	4.162	0	0	0	0	0	0
ZINC05331228	-4.830	-4.964	1.092	-27.357	4.162	0	0	0	0	0	0
ZINC00297848	-4.447	-5.488	1.249	-25.830	4.154	0	0	0	0	0	0
ZINC00282395	-4.829	-5.628	1.028	-22.862	4.146	0	0	0	0	0	0
ZINC59679309	-4.124	-4.991	1.048	-24.546	4.137	0	0	0	0	0	0
ZINC59946761	-5.097	-5.012	0.752	-26.026	4.123	0	0	0	0	0	0
ZINC13125523	-4.846	-5.574	1.792	-25.430	4.121	0	0	0	0	0	0
ZINC16687806	-4.350	-5.697	0.963	-28.394	4.046	0	0	0	0	0	0
ZINC17326384	-4.747	-5.066	1.872	-25.670	4.035	0	0	0	0	0	0
ZINC08683932	-4.063	-5.383	1.256	-26.894	4.034	0	0	0	0	0	0
ZINC59457052	-4.789	-5.147	1.913	-26.098	4.034	0	0	0	0	0	0
ZINC08616455	-4.176	-5.060	1.406	-29.818	4.019	0	0	0	0	0	0
ZINC59302736	-4.579	-4.990	0.763	-25.378	4.009	0	0	0	0	0	0
ZINC03109245	-4.977	-5.016	1.606	-22.534	3.975	0	0	0	0	0	0
ZINC00179512	-4.939	-5.318	1.191	-22.340	3.966	0	0	0	0	0	0
ZINC59514173	-4.244	-5.144	1.994	-25.651	3.896	0	0	0	0	0	0
ZINC00065981	-4.727	-5.290	2.855	-16.643	3.892	0	0	0	0	0	0
ZINC40570819	-4.030	-5.585	1.201	-30.059	3.887	0	0	0	0	0	0
ZINC40570703	-4.221	-5.503	1.427	-24.109	3.880	0	0	0	0	0	0
ZINC59518484	-4.672	-5.141	0.864	-26.180	3.871	0	0	0	0	0	0
ZINC38805548	-4.287	-4.950	1.542	-25.212	3.868	0	0	0	0	0	0
ZINC33594294	-3.840	-4.947	1.129	-25.036	3.865	0	0	0	0	0	0
ZINC06417874	-3.757	-5.402	1.018	-29.516	3.843	0	0	0	0	0	0
ZINC01709000	-4.602	-5.583	1.798	-27.787	3.833	0	0	0	0	0	0
ZINC75278567	-5.092	-5.223	1.593	-23.487	3.829	0	0	0	0	0	0
ZINC16929290	-3.702	-4.963	1.110	-27.311	3.826	0	0	0	0	0	0
ZINC05536639	-4.968	-5.441	1.336	-27.971	3.819	0	0	0	0	0	0
ZINC12495215	-4.462	-5.323	1.300	-21.981	3.817	0	0	0	0	0	0
ZINC00065733	-3.754	-5.045	0.985	-22.200	3.813	0	0	0	0	0	0
ZINC05977376	-4.205	-4.946	1.948	-29.969	3.808	0	0	0	0	0	0
ZINC00047179	-4.999	-5.456	1.482	-26.354	3.798	0	0	0	0	0	0
ZINC00096041	-4.842	-5.370	1.476	-26.751	3.794	0	0	0	0	0	0
ZINC00079148	-3.694	-5.214	1.214	-24.830	3.792	0	0	0	0	0	0
ZINC59508020	-4.189	-5.269	1.287	-27.176	3.788	0	0	0	0	0	0
ZINC14983237	-4.034	-5.212	1.135	-27.155	3.780	0	0	0	0	0	0
ZINC40570829	-3.588	-5.030	0.919	-26.878	3.768	0	0	0	0	0	0
ZINC15973646	-3.564	-5.550	1.289	-24.532	3.721	0	0	0	0	0	0
ZINC03183544	-4.769	-5.144	1.919	-23.297	3.698	0	0	0	0	0	0
ZINC75251431	-5.054	-5.216	1.045	-28.353	3.694	0	0	0	0	0	0
ZINC59519274	-4.591	-5.149	1.165	-24.251	3.689	0	0	0	0	0	0
ZINC01233230	-4.269	-5.259	1.834	-26.959	3.685	0	0	0	0	0	0
ZINC17042267	-4.977	-5.003	1.350	-26.267	3.666	0	0	0	0	0	0
ZINC38976989	-4.711	-5.271	1.233	-26.264	3.658	0	0	0	0	0	0
ZINC00047190	-3.982	-5.313	0.797	-26.367	3.645	0	0	0	0	0	0
ZINC18191555	-4.134	-5.349	1.552	-27.518	3.644	0	0	0	0	0	0
ZINC40570704	-4.581	-5.501	1.415	-28.249	3.628	0	0	0	0	0	0
ZINC59248130	-4.844	-4.962	0.811	-27.991	3.612	0	0	0	0	0	0
ZINC59678778	-3.770	-4.988	0.824	-28.493	3.601	0	0	0	0	0	0
ZINC00297964	-4.969	-5.471	0.753	-23.557	3.597	0	0	0	0	0	0
ZINC18192593	-4.824	-5.331	1.122	-27.279	3.585	0	0	0	0	0	0
ZINC00096273	-4.365	-5.522	1.916	-28.834	3.582	0	0	0	0	0	0

ID	Glide SP-Score	eHiTS-Score	RMSD	LigX	dock_pKi	SP vote	eHiTS vote	RMSD vote	Lig X vote	pki vote	Final Vote
ZINC00255270	-4.669	-5.036	1.002	-29.897	3.579	0	0	0	0	0	0
ZINC40570775	-3.402	-4.983	1.000	-29.491	3.578	0	0	0	0	0	0
ZINC34070709	-4.283	-5.116	2.098	-29.861	3.576	0	0	0	0	0	0
ZINC03149578	-5.060	-5.079	1.050	-25.602	3.575	0	0	0	0	0	0
ZINC12341239	-5.067	-5.204	1.213	-25.500	3.566	0	0	0	0	0	0
ZINC40570706	-4.656	-5.524	1.228	-28.714	3.564	0	0	0	0	0	0
ZINC04673337	-4.079	-5.008	2.105	-26.561	3.532	0	0	0	0	0	0
ZINC05604499	-4.613	-5.021	0.843	-29.887	3.480	0	0	0	0	0	0
ZINC00065933	-4.891	-5.044	1.237	-26.759	3.454	0	0	0	0	0	0
ZINC59678328	-4.156	-4.993	1.969	-24.491	3.451	0	0	0	0	0	0
ZINC00225845	-4.956	-5.012	1.191	-29.529	3.436	0	0	0	0	0	0
ZINC40570772	-5.067	-4.974	0.853	-27.226	3.434	0	0	0	0	0	0
ZINC00198307	-4.932	-5.127	1.817	-27.005	3.432	0	0	0	0	0	0
ZINC05159007	-4.567	-5.205	1.148	-26.423	3.413	0	0	0	0	0	0
ZINC00297927	-4.386	-5.151	0.970	-24.942	3.409	0	0	0	0	0	0
ZINC59248174	-3.801	-4.990	1.711	-25.725	3.347	0	0	0	0	0	0
ZINC59518432	-3.936	-5.375	1.782	-27.084	3.337	0	0	0	0	0	0
ZINC59456920	-4.805	-5.277	1.867	-29.192	3.328	0	0	0	0	0	0
ZINC17201147	-4.117	-5.055	1.990	-26.965	3.304	0	0	0	0	0	0
ZINC17355178	-4.822	-4.953	0.876	-27.919	3.296	0	0	0	0	0	0
ZINC16512576	-4.749	-5.672	1.663	-27.710	3.292	0	0	0	0	0	0
ZINC01087149	-4.305	-5.004	1.209	-26.638	3.286	0	0	0	0	0	0
ZINC00506217	-4.754	-5.201	0.868	-29.841	3.198	0	0	0	0	0	0
ZINC59249700	-4.509	-4.988	1.173	-27.445	3.173	0	0	0	0	0	0
ZINC02043018	-4.747	-4.985	0.813	-28.492	3.152	0	0	0	0	0	0
ZINC00297751	-3.526	-5.353	1.037	-26.479	3.120	0	0	0	0	0	0
ZINC05015728	-4.203	-5.204	1.493	-26.243	3.089	0	0	0	0	0	0
ZINC00943141	-5.065	-5.146	1.609	-27.271	3.087	0	0	0	0	0	0
ZINC04981614	-4.338	-4.986	2.986	-24.856	3.062	0	0	0	0	0	0
ZINC48329990	-3.957	-5.089	1.512	-26.979	3.025	0	0	0	0	0	0
ZINC00062768	-4.438	-5.355	1.800	-22.868	2.982	0	0	0	0	0	0
ZINC40570773	-4.933	-4.966	1.046	-26.633	2.971	0	0	0	0	0	0
ZINC03589752	-4.814	-5.158	0.975	-29.576	2.937	0	0	0	0	0	0
ZINC01274759	-4.641	-5.571	1.186	-27.978	2.869	0	0	0	0	0	0
ZINC00257158	-4.526	-5.280	0.962	-28.263	2.847	0	0	0	0	0	0
ZINC03589755	-4.923	-5.124	1.113	-22.085	2.828	0	0	0	0	0	0
ZINC00297232	-4.730	-5.023	1.166	-28.563	2.727	0	0	0	0	0	0
ZINC06396396	-3.905	-5.427	0.949	-27.015	2.713	0	0	0	0	0	0
ZINC12970788	-4.976	-5.126	1.745	-29.417	2.649	0	0	0	0	0	0
ZINC19772510	-4.470	-5.125	1.096	-26.337	2.279	0	0	0	0	0	0
ZINC00298291	-4.074	-5.034	0.953	-29.960	1.990	0	0	0	0	0	0
ZINC59518301	-4.916	-5.140	1.472	-26.969	1.575	0	0	0	0	0	0

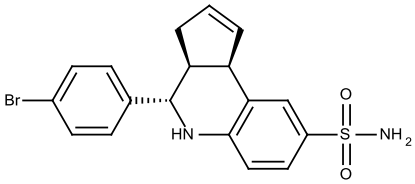
Supplementary table 4.2. Chemical structures of previously reported BF3 inhibitors by Estenabez-Perpina *et al*, Lack *et al*, Munuganti *et al* and commercial anti-androgens.

T3	TRIAC	FLUF
		
9002	9114	9117
		
9145	0056	4035
		
0098	0209	4026
		

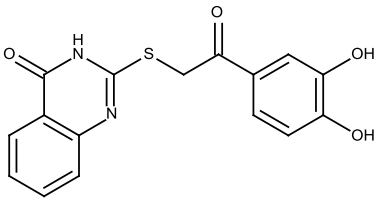
4090



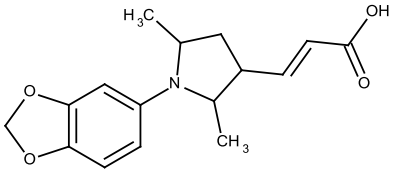
4041



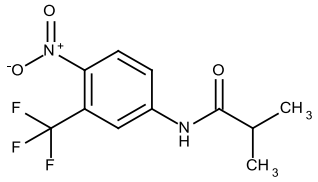
0017



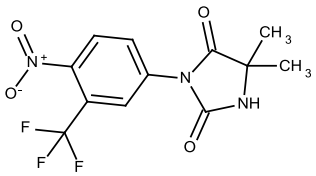
0058



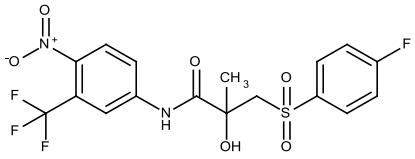
Flutamide



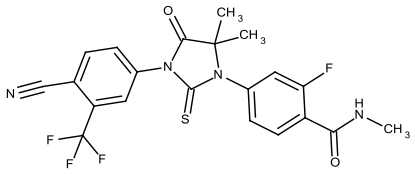
Nilutamide



Bicalutamide



Enzalutamide



Supplementary table 4.3. Lipophilicity-corrected ligand efficiencies (LLE) for the BF3 AR binders.

LLE can be defined as LLE = pIC_{50}-*c*log*P*

ID	eGFP Activity (μM)	pIC_{50}	LogP(o/w)	LLE
13163	0.310	6.509	3.734	2.775
13127	0.600	6.222	3.610	2.612

ID	eGFP Activity (μM)	pIC_{50}	LogP(o/w)	LLE
13167	1.910	5.719	3.323	2.396
13256	0.380	6.420	4.278	2.142
13164	1.200	5.921	4.002	1.919
13255	0.520	6.284	4.868	1.416
13259	0.310	6.509	5.211	1.298
13225	0.700	6.155	4.909	1.246
13221	15.000	4.824	3.648	1.176
13300	9.330	5.030	3.860	1.170
13254	1.200	5.921	4.870	1.051
9099	7.400	5.131	4.123	1.008
13299	8.160	5.088	4.202	0.886
13257	4.180	5.379	4.576	0.803
13235	2.210	5.656	4.876	0.780
13320	25.230	4.598	3.897	0.701
13303	4.460	5.351	4.876	0.475
13310	10.310	4.987	4.535	0.452
13166	1.690	5.772	5.379	0.393
13304	7.270	5.138	5.199	-0.061
13345	13.890	4.857	5.084	-0.227
13321	25.000	4.602	5.123	-0.521

ID	eGFP Activity (μM)	pIC_{50}	LogP(o/w)	LLE
13309	27.210	4.565	5.535	-0.970
13258	6.780	5.169	6.236	-1.067

Supplementary table 5.1. The list of QSAR models built based on distance-dependent descriptors classified as per atomic geometries described in MOE. Desc - Descriptors, PPV - positive prediction values.

Metrics	Desc	Model	ROC	PPV	Specificity	Sensitivity	Accuracy
3	10	Random Forest	0.765	0.806	0.684	0.735	0.717
1	10	Decorate	0.761	0.726	0.646	0.776	0.717
2	10	Bagging	0.759	0.758	0.625	0.712	0.679
1	10	Multilayer-Perception	0.757	0.758	0.667	0.77	0.726
1	10	Random Forest	0.733	0.774	0.659	0.738	0.708
3	10	Decorate	0.715	0.677	0.623	0.792	0.708
1	15	ADTree	0.707	0.726	0.638	0.763	0.708
1	15	Decorate	0.787	0.758	0.688	0.81	0.755
2	15	Random Forest	0.767	0.806	0.684	0.735	0.717
1	20	Multilayer-Percepiron	0.763	0.774	0.689	0.787	0.745
2	20	ADTree	0.739	0.79	0.683	0.754	0.726
1	20	Logit Boost	0.738	0.758	0.651	0.746	0.708
3	20	ADTree	0.737	0.79	0.683	0.754	0.726
2	20	Logit Boost	0.717	0.758	0.643	0.734	0.698
3	20	Decorate	0.714	0.742	0.61	0.708	0.67
3	20	Logit Boost	0.718	0.758	0.643	0.734	0.698
Average			0.742	0.759	0.657	0.753	0.712