

**THE USE OF CHEMINFORMATICS METHODS FOR PREDICTING ADVERSE
DRUG RESPONSES BY HUMAN ANDROGEN RECEPTOR**

by

Naman Paul

B. Tech. (Hons) Bioinformatics, Amity University Rajasthan, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Bioinformatics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2016

© Naman Paul, 2016

Abstract

The human Androgen Receptor (AR) is a ligand-activated transcription factor that plays a pivotal role in the development and progression of prostate cancer (PCa). AR is also critical for the survival of many forms of castration resistant prostate cancer (CRPC). The currently used AR inhibitors (anti-androgens) face clinical limitations as drug resistance has been reported in patients, both primary and acquired. In 20% of the CRPC patients resistance to AR antagonists arise due to the mutations in the androgen binding site (ABS) of the receptor. Some mutations can convert antagonist to agonist. Such gain-of-function mutations have been reported across the length of the ligand binding domain (LBD) of AR that contains the ABS, it is imperative to develop a prognostic personalized therapy platform which would equip clinicians with actionable strategies in regard to previously unreported AR aberrations when they are encountered in clinical samples. The goal of this study is to develop a theoretical approach that can characterize such previously unreported AR mutants and predict their response to the currently used anti-androgens.

Thus, a novel ‘in-silico’ pipeline has been created that amalgamates the state-of-the-art cheminformatics methods with experimental assays that enable predicting AR mutants and characterizing their drug responses with high accuracy. The corresponding pipeline utilizes QSAR approach that extracts key protein-ligand interactions quantified by the in-house developed 4D-inductive molecular descriptors. The developed QSAR models reach about 90% accuracy that forecasts agonist or antagonist behaviors of AR mutants caused by clinically used and experimental anti-androgens. Furthermore, a previously unreported mutant, T878G has been predicted to be activated by both first and second generation anti-androgens and the corresponding experimental evaluation confirmed this prediction. Finally, the applicability and adaptability of the developed cheminformatics pipeline was tested against an experimental anti-androgen drug ODM-201 which was not a part of the QSAR training dataset, and the predictions were confirmed by experimental evaluations. Overall, the developed pipeline can provide useful insights towards understanding the changing genomic landscape of advanced PCa.

Preface

The project idea was conceived by Dr. Artem Cherkasov. This work consists of my contributions towards the development of a personalized prognostic platform to predict and monitor mutations in patients.

A version of the work described in Chapters 2, 3 and 4 has been published, [Paul N, Carabet LA, Lallous N, Yamazaki T, Gleave ME, Rennie PS, Cherkasov A. *Cheminformatics Modeling of Adverse Drug Responses by Clinically Relevant Mutants of Human Androgen Receptor*. J. Chem. Inf. Mod., November 2016, DOI: 10.1021/acs.jcim.6b00400]. Drs. Cherkasov and Rennie are the senior authors and have supervised this project as well as manuscript revision. I performed all of the computational experiments, as well as drafted and revised the manuscript. The SVL script for 4D-Inductive descriptors has been created by Carabet LA. Dr. Lallous N, performed the biological evaluation of the predicted mutants. Parts of the methods and results described in Chapters 2, 3 and 4 have also been described in the aforementioned publication. The chemical structure of ODM-201 has not been shown.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	iv
List of Tables	viii
List of Figures.....	ix
List of Symbols	xii
List of Abbreviations	xiii
Acknowledgements	xvi
Dedication	xvii
Chapter 1: Introduction	1
1.1 Prostate cancer and androgen receptor targeted therapy.....	2
1.1.1 Role of androgen receptor in prostate cancer	4
1.1.2 Castration resistant prostate cancer (CRPC).....	7
1.1.3 AR mutants and therapy resistance.....	8
1.2 Computer-aided drug design.....	10
1.2.1 Structure-based methods.....	10
1.2.1.1 Molecular Docking	11
1.2.1.2 Molecular dynamics simulations	12
1.2.2 Ligand-based methods	13
1.2.2.1 Quantitative structure activity relationship (QSAR) modeling	13
1.2.2.2 Evaluating QSAR models	14
1.3 Cheminformatics.....	16

1.3.1	Machine learning in Cheminformatics and Classification algorithms used	16
1.3.1.1	DecisionStump.....	17
1.3.1.2	OneR.....	17
1.3.1.3	RandomForest.....	18
1.3.1.4	Bagging.....	19
1.3.1.5	Dagging.....	20
1.3.1.6	IBk.....	20
1.3.1.7	LibSVM	21
1.3.2	Exploratory data analysis for attribute selection.....	22
Chapter 2: Cheminformatics platform development		24
2.1	Datasets.....	27
2.1.1	Training dataset.....	27
2.1.2	Test dataset.....	29
2.1.3	Ligand molecules used.....	29
2.2	Molecular docking protocol.....	30
2.2.1	Protein structure preparation.....	30
2.2.2	Receptor grid generation.....	31
2.2.3	Glide XP docking mode.....	31
2.3	Molecular descriptor computation	31
2.3.1	Glide per-residue energy scores.....	32
2.3.2	4D-Inductive descriptors.....	33
2.3.3	Attribute pruning and prioritization.....	35
2.3.4	Attribute selection.....	36

2.4	QSAR model development	37
2.4.1	Model building and consensus vote approach	37
2.4.2	Evaluation of QSAR models.....	39
2.4.3	External test set validation	41
2.5	Molecular dynamics simulations analyses.....	41
2.5.1	Steps involved in MD simulations.....	41
2.5.1.1	Ligand geometry optimization and atomic charge assignment.....	42
2.5.1.2	Protein-ligand forcefield assignment and energy minimization	42
2.5.1.3	Heat, pressure application and production run	43
2.5.2	RMSD and contact frequency analysis	43
2.6	In-vitro screening of predicted anti-androgen responses	44
Chapter 3: Cheminformatics modeling of AR mutant— drug responses		45
3.1	In-silico AR mutant analyses	45
3.1.1	Generating AR mutant structures.....	45
3.1.2	Structure-based analyses	46
3.2	Molecular descriptors and statistical trends	48
3.2.1	List of 4D-inductive descriptors computed.....	49
3.2.2	Drug response correlation with molecular descriptor values.....	53
3.2.3	Descriptor pruning and ranking	54
3.3	QSAR model development and validation	57
3.3.1	Categorical QSAR models predict biological responses	59
3.3.2	Significance of consensus voting for classification algorithms	60
3.3.3	Statistical accuracy of QSAR predictions.....	60

3.3.4	Applicability domain assessment.....	62
3.3.5	Predicted mutants.....	64
3.4	Molecular dynamics simulations analysis	65
3.4.1	Protein structural stability evaluation	66
3.4.2	Contact frequency analysis	68
3.5	In-silico evaluation of ODM-201 – AR mutant responses	69
Chapter 4: T878G mutant agonizes anti-androgens		71
4.1	Structural analysis of T878G – ligand interactions.....	71
4.2	Descriptor value correlation of T878G – anti-androgen complexes.....	75
4.3	Predicted and experimental biological responses	76
4.4	MD simulations analysis of T878G – anti-androgen complexes.....	76
4.4.1	RMSD analysis determines structural stability.....	77
4.4.2	Contact frequency analysis of T878G – anti-androgen complexes	80
4.5	Experimental evaluation of the T878G mutant response to anti-androgens.....	81
4.6	ODM-201 is ‘effective’ against the T878G mutant.....	83
Chapter 5: Conclusions		85
5.1	Summary of the study	85
5.2	Novel mutants predicted	86
5.3	Importance of drug response characterization and future scope.....	87
Bibliography		88
Appendices.....		100
	Appendix A QSAR modeling dataset.....	100
	Appendix B List of 225 attributes screened for importance by Boruta	101

List of Tables

Table 1.1: Treatment options for PCa management	3
Table 2.1: List of clinically reported AR wild-type substitutions	28
Table 2.2: Ligand molecules used for docking, containing native ligand DHT and other anti-androgens	30
Table 2.3: Different categories of molecular descriptors used in QSAR modeling	32
Table 3.1: Residues of the wildtype AR (WT-AR) mutated to engineer in-silico mutants.....	46
Table 3.2: List of 4D-inductive descriptors computed, AA# descriptors were calculated for the mutated residues.....	53
Table 3.3: Descriptor value correlation with biological response obtained for mutant H875Y ...	54
Table 3.4: List of molecular descriptors with confirmed importance as per Boruta implementation for attribute prioritization.....	57
Table 3.5: Attributes selected for QSAR model construction	59
Table 3.6: Performance statistics of QSAR classification models	61
Table 3.7: AR-LBD mutants predicted to yield an agonist response towards anti-androgens	64
Table 3.8: RMSD comparison between initial docked structures and energy equilibrated MD structures	68
Table 4.1: Descriptor value correlation with predicted biological response for mutant T878G ..	76
Table 4.2: Predicted and experimental biological responses of T878G mutant to DHT and anti-androgens	76
Table 4.3: RMSD comparison between initial docked and energy equilibrated structures.....	77

List of Figures

Figure 1.1: Location of AR gene at q11-12 of the X Chromosome, Exon 1 encodes for N-terminal domain (NTD) of the full length AR protein. Exons 2-3 encode DNA binding domain (DBD); exon 4 encodes Hinge (H) region. Exons 5-8 encode ligand binding domain (LBD).....	4
Figure 1.2: Native ligand DHT bound to the androgen binding site (ABS) of AR, key interacting protein residues have been shown in cyan such as Asn 706, Met 746 and Thr 878.....	7
Figure 1.3: Enzalutamide binding differently to mutant AR, F877L (left) and wild-type receptor (right)	9
Figure 1.4: A general QSAR workflow, the dataset is generated followed by statistical analyses and evaluation followed by experimental verification to predict activity/inactivity of a molecule	14
Figure 1.5: Workflow of DecisionStump algorithm, classifying based on a 1-level decision tree	17
Figure 1.6: One-rule based classifier algorithm, selects the rule for classification with minimum error frequency.....	18
Figure 1.7: RandomForest algorithm implementation.....	18
Figure 1.8: Classification using Bagging algorithm workflow, splits the dataset into smaller subsets which is supplied to classifier followed by majority voting	19
Figure 1.9: Dagging algorithm workflow	20
Figure 1.10: Classification using k nearest approach (IBk), shown is the classification of a test set instance when k=3	21
Figure 1.11: LibSVM algorithm workflow for classification.....	22
Figure 2.1: Initial pipeline used for predicted mutant - anti-androgen responses	25

Figure 2.2: Final functional pipeline developed for studying mutant- anti-androgen responses .	26
Figure 2.3: WEKA knowledge workflow implemented	38
Figure 3.1: Ligand interaction diagram-Enzalutamide bound to the androgen binding site (ABS) of F877L mutant, L877 residue interacts with the ligand molecule through H-bonding in addition to π - π stacking between benzene ring of F765 and ring moiety of Enzalutamide.....	48
Figure 3.2: Prioritization of attributes, initial run of Boruta package. Shadow attributes: blue, Rejected attributes: red, Undecided importance of attributes: yellow, Confirmed importance: green.....	55
Figure 3.3: Attribute prioritization, final run- segregates into confirmed (green) and rejected (red) classes based on importance	56
Figure 3.4: Determining the number of attributes to be used for QSAR modeling.....	58
Figure 3.5: Over 85% of area under the ROC curve, demonstrates the high diagnostic ability, effectively discriminating true positives from the false positives	62
Figure 3.6: Applicability domain assessment of molecular descriptor values, utilized in QSAR modeling. Red arcs represent the spread of the Test set values, which are within the range of the Training set values (Green arcs) \pm 15%.....	63
Figure 3.7: Structural stability of the WT AR - Enzalutamide complex illustrated through a stable trajectory obtained for 25ns MD simulation.....	67
Figure 3.8: F877L/T878A - Enzalutamide complex trajectory obtained upon 25ns MD simulation	68
Figure 3.9: Contact frequency analysis of F877L/T878A - Enzalutamide complex	69
Figure 4.1: Bicalutamide, Enzalutamide and Hydroxyflutamide bound to ABS of T878G mutant	71

Figure 4.2: T878G mutant activated by DHT, H-bond formation with L874 residue.....	72
Figure 4.3: Bicalutamide interacts with the T878G ABS pocket through π - π stacking interactions with benzene rings of F892, W742, F765 and H-Bond interaction with N706 sidechain and backbone of L705	73
Figure 4.4: Enzalutamide interacts with the T878G mutant pocket by H-bond formation with G878 and π - π stacking interactions between the benzene rings of F765 and F892.....	73
Figure 4.5: Enzalutamide bound to T878G ABS, benzene rings of F892 and F765 interact through T-shaped π - π stacking with <i>a</i> and <i>b</i> rings of Enzalutamide	74
Figure 4.6: Hydroxyflutamide interacting with the T878G mutant ABS pocket through H-bonding with L705 backbone and sidechain of N706, also is seen a π - π stacking interaction with benzene ring of F765	75
Figure 4.7: Stable trajectory obtained for T878G receptor - Enzalutamide complex.....	78
Figure 4.8: RMSD analysis of T878G-Bicalutamide complex.....	79
Figure 4.9: RMSD analysis of T878G-Hydroxyflutamide complex	80
Figure 4.10: No major alterations seen in the MD trajectory of Hydroxyflutamide bound T87G mutant ABS pocket.....	80
Figure 4.11: Contact frequency comparison for T878G-Enzalutamide complex.....	81
Figure 4.12: The response of T878G mutant to Enzalutamide, Hydroxyflutamide, and Bicalutamide, in an in vitro cell-based assay. Each concentration was assayed in quadruplicate n = 4, with a biological replicate of n = 2.	82
Figure 4.13: Stable trajectory of the T878G pocket bound to ODM-201.....	83
Figure 4.14: ODM-201 yields antagonist response towards T878G mutant	84

List of Symbols

\AA Angstrom

Σ Summation

π Pi

\emptyset Biological activity as a function of molecular descriptors

ATM Standard atmospheric pressure

K Kelvin

kcal Kilo calories

K_D Equilibrium dissociation constant

IC₅₀ Inhibition at 50% concentration

List of Abbreviations

3D	3 Dimensional
4D	4 Dimensional
ABS	Androgen Binding Site
ADT	Androgen Deprivation Therapy
ADMET	Absorption Distribution Metabolism Excretion Toxicity
AMBER	Assisted Model Building with Energy Refinement
AR	Androgen Receptor
AUC	Area Under Curve
BPH	Benign Prostatic Hyperplasia
CADD	Computer Aided Drug Discovery
caret	classification and regression tools
CRPC	Castration Resistant Prostate Cancer
DBD	DNA Binding Domain
DHT	5 α - Dihydrotestosterone
DRE	Digital Rectal Examination
ECHA	European Chemicals Agency
EPA	United States Environmental Protection Agency
FDA	United States Food and Drug Administration
FN	False Negative
FP	False Positive

FPR False Positive Rate

GAFF General Amber Force Field

GOLD Genetic Optimization for Ligand Docking

HPCC High Performance Computing Cluster

IBk Instance Based k

ICM-dock Internal Coordinate Mechanics dock

LBD Ligand Binding Domain

LibSVM Library for Support Vector Machines

MB Megabytes

MD Molecular Dynamics

MZSA Maximum Z Score Attribute

NMR Nuclear Magnetic Resonance

NTD N-Terminal Domain

OneR One Rule

PCa Prostate Cancer

PSA Prostate Specific Antigen

QSAR Quantitative Structure-Activity Relationship

RESP Restrained electrostatic potential

RFE Recursive Feature Elimination

RMSD Root Mean Square Deviation

ROC Receiver Operating Characteristic

SVL Scientific Vector Language

TN True Negative

TP True Positive

VMD Visual Molecular Dynamics

WEKA Waikato Environment for Knowledge Analysis

Acknowledgements

I would like to thank my supervisor, **Dr. Artem Cherkasov** for his tremendous mentorship and guidance, and giving me an amazing opportunity to work with few of the best minds. His words of encouragement and appreciation catalyzed my research project, timely critiques helped me shape the project work with precision imperatively required for keeping on the right track, maintaining pace ultimately leading to the successful completion of my project. I cannot sufficiently express my gratitude and admiration of his approach to dealing with novel concepts, providing me the liberty to experiment and learn from both successful and unsuccessful outcomes.

I would like to thank my supervisory committee members, **Dr. Paul S. Rennie, Dr. Alexander W. Wyatt and Dr. Colin Collins**, for their time, valuable insights and motivation; and **Dr. Inanc Birol** for chairing the examination committee. I further extend my gratitude to the faculty members from both UBC and SFU, who taught various courses during the Bioinformatics graduate program helping me develop new analytical skills. I would like to especially thank, the Bioinformatics program coordinator **Sharon Ruschkowski** for being the very first contact person, for helping me through the application process, and supporting throughout with her prompt responses. I would also like to express my gratitude to the director of the Bioinformatics graduate program, **Dr. Steven Jones** for his support through all stages.

I thank all my lab members, who have always stepped up to reach out whenever I needed any assistance. I thoroughly enjoyed all the interactions (both scientific and non-scientific) with them. A special shout out to **Dr. Nada Lallous** who was involved in the experimental testing of the predicted mutants. I can't thank enough **Dr. Takeshi Yamazaki** for his expert guidance on molecular dynamics simulations and for patiently answering all my queries.

No words can help me express my gratitude for my parents and sister for believing in me and for always being present as my constant support system; for their unconditional love, and prayers that strengthened me to live in Canada, miles away from my home in India. I thank the God Almighty, for all His abundant blessings bestowed upon me, being the light of my path at every moment. In the social paradigm, I would like to thank the members of **St. Anselm's Anglican Church** for accepting me into the parish and being my big Canadian family.

Last but not the least, I would take this opportunity to thank my friends and colleagues, who stood by me through streams and storms; my teachers from **Amity University Rajasthan** and my alma mater, **St. Mary's Convent Senior Secondary School, Dewas** for inspiring me to live by the motto "Let your light shine".

To my parents, Nalini and Alok

my sister, Anubhuti

Chapter 1: Introduction

Prostate cancer (PCa) is one of the most commonly diagnosed malignances; it is estimated that on average 1 in 8 Canadian men will be diagnosed with PCa, and more than 21,000 new PCa incidences will be accounted every year.¹ According to a Statistics Canada report (released in April 2016), in the year 2012, PCa claimed over 3700 lives, and this number is estimated to increase up to 4000 in the year 2016.² It is also, the second leading cause of cancer-related fatalities in North American men.³ A majority of the patients are diagnosed in early stages, when the disease is localized. However, approximately 40% of the patients progress into more aggressive and invasive forms.⁴ These aggressive forms of PCa are often associated with treatment failure, due to drug resistance development towards primary treatment. One of the mechanisms associated with treatment failure is the emergence of mutations in the human androgen receptor (AR) gene. This has been observed in approximately 18-20% of all advanced PCa cases.⁵ Anti-androgens are used to control the cell proliferation that act as antagonists, by binding to the androgen binding site (ABS) but not activating the AR receptor.⁶ However, certain somatic mutations are ascribed to drug resistance, which alter the response to anti-androgens from antagonist to agonist leading to receptor activation.⁷ The functional characterization of these mutants revealed such gain-of-function scenarios that signify the importance of identifying such mutations and monitoring the therapeutic response in patients.⁸

With the help of this study, we aim to investigate the AR receptor mutations and to identify new ones responsible for anti-androgen treatment failure in order to build up a prognostic platform using evidence-based approaches. An amalgamation of modern computer-aided drug discovery (CADD) methods, cheminformatics and experimental validation has been employed to achieve

the goal of modeling previously unreported AR mutants and to predict their response to a panel of clinically used anti-androgens.⁹

1.1 Prostate cancer and androgen receptor targeted therapy

The human prostate is a walnut sized gland comprising of a median and two lateral lobes, located between the urinary bladder and the penis.¹⁰ The most important function of the prostate is secretion of nutritional components that liquefy the coagulated semen, nourishes as well as protects it. The prostate is also involved in controlling micturition, as the muscle fibers surround the urethra shrink to slow down and stop the urine flow.¹¹ Medical conditions associated with the prostate are the following:

- Prostatitis - It is the tenderness of the prostate caused due to an infection. This condition can be well managed via antibiotics.¹²
- Benign prostatic hypertrophy (BPH) – It is the enlargement of the prostate, obstructing the normal passage of urine. Urination becomes a daunting task, and delay in treatment/management could lead to serious consequences such as an urgent bladder emptying procedure.¹³
- Prostate cancer (PCa) – The uncontrolled proliferation of cells of the prostate is termed as Prostate cancer (PCa).¹⁴ This diagnosis is usually performed by digital rectal examination (DRE) to determine any unusual protrusions and lumps. This is followed up by a blood screening to quantify prostate specific antigen (PSA) levels.¹⁵ Generally, higher levels of PSA are observed in enlarged prostates. The other confirmatory tests are prostate biopsy, and prostate ultrasound examination.¹⁶ PCa is characterized by a complex range of factors such as race, lifestyle, familial history, age, nutrition, overall health status etc.¹⁷

There are several treatment options available for PCa management. Some of these are listed in Table 1.1 below.

<i>S. No.</i>	<i>Treatment option</i>	<i>Description</i>
1	Active surveillance	Close monitoring the status of the prostate, slowly progressing disease ¹⁸
2	Prostatectomy	Removal of the prostate ¹⁹
3	Radiation therapy	Bombarding the tumor cells with high intensity radiation ²⁰
4	Hormone therapy	Also called androgen deprivation therapy (ADT), cutting off androgen supply to cancerous cells to inhibit cell growth ²¹
5	Chemotherapy	Administration of anti-cancer drugs, affects both healthy and malignant cells ²²

Table 1.1: Treatment options for PCa management

1.1.1 Role of androgen receptor in prostate cancer

The human androgen receptor (AR) is one of the proteins that plays a pivotal role in the development and progression of PCa.²³ AR is over expressed in PCa, linked to growth of the prostate cells, and is activated by binding of androgen steroids, such as 5 α - Dihydrotestosterone (DHT) and testosterone, to the Androgen binding site (ABS) of the receptor.²⁴ This interaction results in the nuclear translocation of the protein dimer, where its DNA response elements bind to the DNA, thus transcriptionally activating the AR.²⁵ The AR gene is located in q11-12 region of the X chromosome.²⁶ Full length AR protein is made up of 919 amino acids, and is organized into several structural and functional domains.²⁷ (See Figure 1.1)

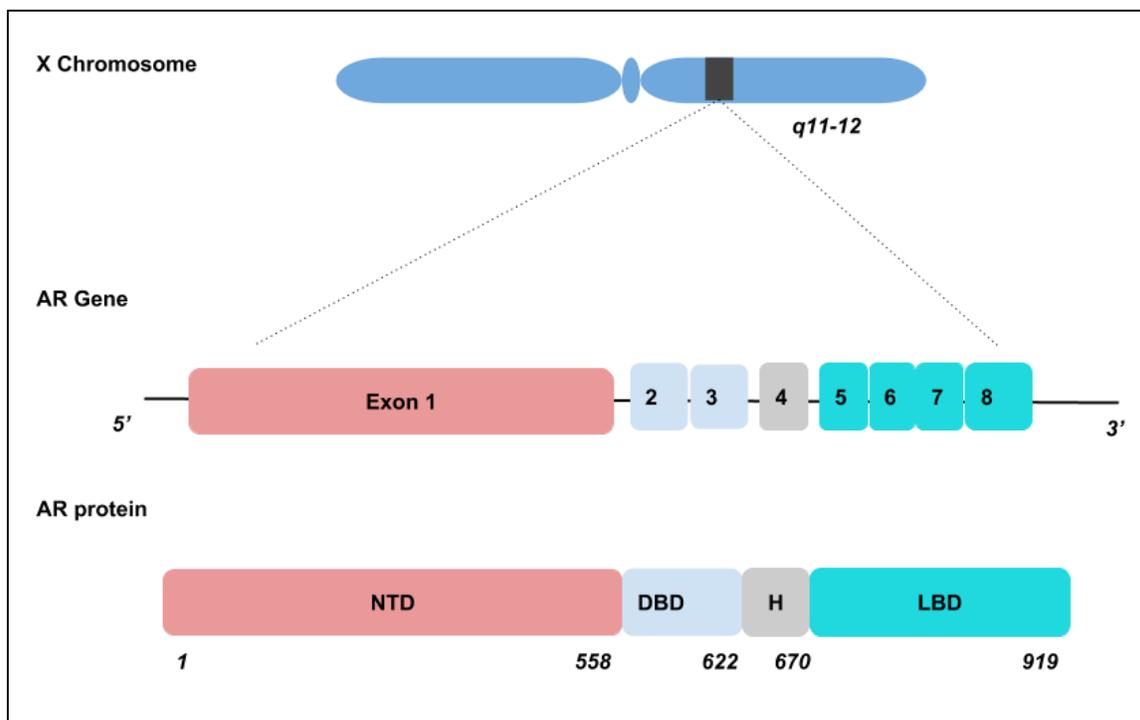


Figure 1.1: Location of AR gene at q11-12 of the X Chromosome, Exon 1 encodes for N-terminal domain (NTD) of the full length AR protein. Exons 2-3 encode DNA binding domain (DBD); exon 4 encodes Hinge (H) region. Exons 5-8 encode ligand binding domain (LBD).

N-terminal domain (NTD):

The N-terminal domain (NTD) makes up about 60% of the AR.²⁸ This is the largest AR domain encoded by exon 1, comprising of 558 amino acid residues (AR residues # 1-558). This is the least conserved domain. NTD usually contains repeated sequences of Glycine (G) and Glutamine (Q). The presence of more than 35 Q repeats is associated with muscular atrophy/ Kennedy's disease.²⁹ The AR NTD is involved in controlling the AR transcriptional activity by recruitment of various transcriptional machinery components. It also contains the FxxLF and WxxLF motifs that are involved in intra-molecular interaction with the ligand binding domain (LBD).³⁰⁻³¹ NTD most importantly contains the transcriptional regulatory AF-1 region (AR residues #142- 485).³²

DNA binding domain (DBD):

The DNA binding domain (DBD) is encoded by exons 2 and 3. DBD (AR residues # 559-622) contains the androgen response elements (ARE) that interact with the DNA.³³ It is highly conserved and bears high sequential identity to other nuclear receptors- estrogen receptor (ER), glucocorticoid receptor (GR) and progesterone receptor (PR).³⁴ DBD consists of 2 Zinc finger structures where the 2 Zinc atoms are held by tetrahedral coordination with C.³⁵

Hinge region (H):

The hinge region is highly flexible and connects the DBD and LBD (AR residues # 623-670). It is involved the regulation of receptor translocation, nuclear transportation and DNA selectivity.³⁶

Ligand binding domain (LBD):

The ligand binding domain (LBD) is one of the most well studied and characterized structural domains of the AR that makes up of about 1/4th of the total length (AR residues # 671-919).²⁷ AR-LBD can exist on its own, independent of the full length AR, making it easier to biosynthesize and experimentally test.³⁷ It encloses the ABS, which is the primary target site for

androgen binding. Androgens such as 5α -DHT, bind into the hydrophobic ABS pocket formed by α -helical folds.³⁸ (See Figure 1.2) The binding of androgens lead to induction of structural changes that aid the translocation of the AR into the nucleus.³⁹ Numerous structures of AR LBD have been solved through X-ray crystallography.⁴⁰ The AR LBD structure can be described as a three-layered helical sandwich composed of 11 α -helices and 2 sheets made by 4 β -strands. Much has been previously described about the ‘lid’ or agonist conformation attained by helix 12 which holds the androgen in place with the C-terminal region further contributing to the ligand stabilization mechanism.⁴¹ The ABS also serves as the target site for hormone therapy agents (anti-androgens) administered to PCa patients. Distinct somatic mutations have been identified in CRPC patients, which are located across the AR LBD, making it an important investigatory region.⁴² The anti-androgens in the absence of androgens bind to the ABS with a high affinity, depriving the prostate cells of the androgens which are critical to their survival and development.⁴³

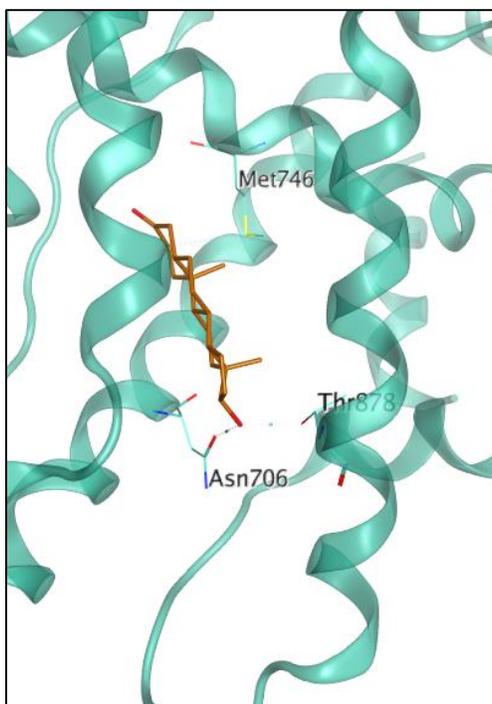


Figure 1.2: Native ligand DHT bound to the androgen binding site (ABS) of AR, key interacting protein residues have been shown in cyan such as Asn 706, Met 746 and Thr 878

1.1.2 Castration resistant prostate cancer (CRPC)

It is estimated that up to 40% of PCa cases will progress into more aggressive metastatic stage.⁴ The Androgen Deprivation Therapy (ADT) is one of the primary treatment options for the treatment of aggressive PCa.⁴⁴ Due to the dependence of AR on androgens, eliminating the supply of male hormones and their replacement with anti-androgens that inhibit regular AR activity result in a decline of the PSA levels.⁴⁵ However, the efficacy of ADT to maintain low PSA concentrations diminishes over time, promoting most of the patients into a metastatic form of the disease called as metastatic castration resistant prostate cancer (CRPC), which usually occurs within < 24 months on ADT.⁴⁶ The CRPC is extremely invasive, and lethal in some cases.⁴⁷ Strategies to treat CRPC are still mostly experimental, and range from the use of new anti-androgens to the introduction of next-generation hormone therapies, however overall CRPC

treatment stands as a difficult challenge.⁴⁸⁻⁵³ Furthermore, in certain cases, patients unresponsive to CRPC treatment can progress into more aggressive forms such as the Neuro Endocrine Prostate Cancer (NEPC).⁵⁴

1.1.3 AR mutants and therapy resistance

In untreated primary prostate cancer the AR is unaltered.⁵⁵ However after ADT, about 10-30% of CRPC patients harbor mutations in their AR.⁵⁶ For instance, a recent seminal study by Robinson and colleagues showed that the AR is altered in 63% of CRPC patients: amplification (52%), mutation (18%).⁵⁵ These aberrations are critical from a clinical perspective, as they are typically associated with therapy resistance and treatment failure. Mutations in AR can change its function. Certain AR mutations can lead to non-specific targeting of the AR, resulting in conversion of clinical anti-androgens from antagonists into agonists, as well as AR activation by lower ligand concentrations.⁵⁷⁻⁵⁸ Furthermore, these mutations result in making AR promiscuous with decreased ligand specificity that enables binding of other steroids such as progesterone, estrogen, glucocorticoids.⁵⁹ Some frequently observed AR mutations (with known agonists) include, L702H (Glucocorticoids), W742L/C (Bicalutamide), H875Y (Progesterone, Hydroxyflutamide, Bicalutamide), F877L (Enzalutamide, ARN-509), T878A (Hydroxyflutamide, Glucocorticoids).^{7, 60}

As described in the previous sections, the AR ABS is the target site for the current anti-androgens which bind with a higher affinity in the absence of the native ligand DHT. The continuous pressure on the ABS by the anti-androgens can result in the occurrence of point mutations in the corresponding DNA sequence encoding the AR-LBD.⁶¹ These acquired single point mutations cause substitution of amino acids, thus leading to overall structural change in the AR ultimately resulting in response alteration in CRPC patients.⁴³ Interestingly, the mutations

have significant structural impacts on the receptor as well as the anti-androgen binding pose and conformation within the ABS pocket. For the F877L mutant, substitution of Phenylalanine (F) by Leucine (L) alters the interactions as an outcome of structural modification.⁶² Molecular modeling reveals that structural modifications permit ligand molecules such as Enzalutamide to traverse further into the ABS pocket and interacts in a distinct manner when compared to the wild-type (WT).⁶³ (See Figure 1.3)

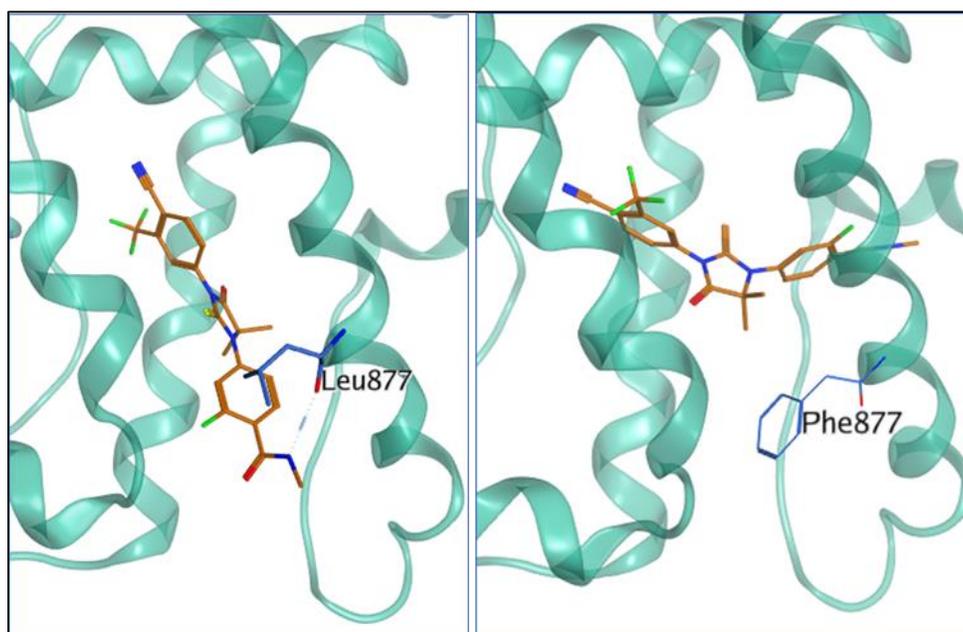


Figure 1.3: Enzalutamide binding differently to mutant AR, F877L (left) and wild-type receptor (right)

The AR aberrations that emerge in CRPC patients can be classified into the following three categories: AR gene amplification, somatic mutations leading to promiscuous AR, splice variants that may result in ligand-independent activation. Our work focusses on somatic AR mutations that can convert anti-androgens from antagonist to agonist.

Recently, our group performed large-scale functional characterization of 24 AR-LBD mutants to study their response to various available anti-androgens, as well as to other steroids such as estradiol, progesterone, and glucocorticoid. Our findings indicate that two of the mutations cause

an increase in the sensitivity of AR to other steroids. It has also been previously shown that several mutants such as the T878A, H875Y etc. can be stimulated by nanomolar progesterone concentrations and activated by estradiol.⁸ This evidently highlights one of the escape mechanisms the AR inhibitors undertake.

1.2 Computer-aided drug design

Computer-Aided Drug Design (CADD) technology has transformed the face of drug discovery and development over the years.⁶⁴ The application of CADD ranges from modeling protein structures to predicting absorption, distribution, metabolism, excretion, toxicology (ADMET) properties of small molecules, thus playing an important role in the drug discovery and development process.⁶⁵ CADD methods can be employed to accomplish complex computations in lesser time reducing both experimental testing and time consumption overheads.⁶⁶ The CADD techniques can be employed to perform modeling, simulations and generate statistical predictions based on the structural, biological, and chemical information available through resources such as databases, experimental results. Various CADD techniques are widely used in pharmaceutical, agrochemical industries as well as by toxicity assessment agencies such as Health Canada, US-FDA (United States Food Drug Administration), ECHA (European Chemicals Agency), US-EPA (United States Environmental Protection Agency) to name a few. The CADD methods can broadly be classified into structure and ligand-based methods.⁶⁷ Those methods that have been employed in this study have been described in greater detail in the subsequent sections.

1.2.1 Structure-based methods

Structure-based CADD methods rely on the availability of molecular structures of biological macromolecules such as proteins.⁶⁷ These methods are oriented around the molecule's

requirement to interact with amino acid residues within a protein's binding site. Numerous 3D structures obtained through X-ray crystallography and NMR spectroscopy can be utilized to investigate and determine target-molecule interactions, speeding up the drug discovery process.⁶⁸ Considering the protein-binding site, very large compound libraries, such as the ZINC¹² containing over 35 million purchasable compounds can be screened in a speedy manner to narrow down a few thousand hits.⁶⁹ This process is termed as virtual screening. A few of the structure-based CADD methods include: Protein homology modeling (modeling a protein structure, based on similarity to a template protein), binding site identification (protein region where small molecules bind).⁷⁰ The methods used in this study were: molecular docking, and molecular dynamics simulations.

1.2.1.1 Molecular Docking

Molecular docking can be defined as a method to predict how two molecules interact with each other, by formation of an intermolecular complex.⁷¹ Traditionally, this involves docking of a small molecule into a binding site of a macromolecule such as protein. It is widely used to model binding poses and conformations of small molecules within the protein pocket.⁷² There are two major types of docking that can be performed: rigid and flexible. The ligand molecule is treated as a rigid entity in Rigid docking. Only translational and rotational degrees of freedom are allowed, that are used to generate large number of conformations which are then separately docked.⁷³ On the other hand, flexible docking involves generation of conformations on the fly, through stochastic search methods with greater degrees of freedom and randomness allowed.⁷⁴ The binding fitness of the small molecule is analyzed within the receptor's binding site.⁷⁵ Some of the most commonly used molecular docking suites are: Glide (Schrödinger), AutoDock, eHiTS, GOLD and ICM-Dock among many others.⁷⁶⁻⁸¹

1.2.1.2 Molecular dynamics simulations

Molecular dynamics (MD) simulations have catalyzed the drug discovery process.⁸² This technique has evolved as an impactful tool due to development of better computation algorithms as well as hardware support.⁸³ With the help of MD simulations the overall flexibility and stability of the protein – ligand (target – drug) system can be assessed.⁸⁴ In classical MD simulations the molecular interaction and motion is studied according to Newtonian physics.⁸⁵ The resultant trajectory or path illustrates how the positions and velocities of the particles in the system vary with time. The obtained trajectory can be represented by the following based on Newton's second law of motion, where F is force, m is mass and a is acceleration:

$$F = ma$$

In biomolecule modeling, force fields are employed for the estimation of energy of the system and forces that govern interatomic crosstalk. The extensive conformational sampling of both the ligand and receptor molecules are performed when placed in a solvated system that is pressurized, heated, and energy equilibrated. As a result, a number of configurations of the protein-ligand system are generated that indicate the trajectories specifying the atomic coordinates, locations and velocities over the simulation time.⁸⁶ Furthermore, other properties such as total energy of the system, kinetic and potential energies can also be calculated. The computation of such energy terms can be used to predict the folding of a protein structure from an initial unfolded state.⁸⁷ MD simulations can also be used to model the ligand binding kinetics starting from a random ligand position, to determine the target binding site.⁸⁸ The comparison between the initial and equilibrated structure can provide atomic level insights of how the system has evolved over the simulation time when exposed to ambient conditions.⁸⁹

1.2.2 Ligand-based methods

The ligand-based methods involve the study of ligand molecules that are known to interact with a protein target of interest. These methods are highly dependent on the knowledge derived from previously known compounds that bind to the same active site of the protein, or might interact with the other members of the same protein family.⁹⁰ These methods can also be used to select small molecules based on their structural similarity, or functional attributes. Furthermore the ligand-based CADD methods can be employed to predict the biological activity of a compound based on previously known compound-activity relationship through Quantitative Structure Activity Relationship (QSAR) modeling.⁹¹

1.2.2.1 Quantitative structure activity relationship (QSAR) modeling

QSAR is a technique used to relate numerical measures and properties (endpoints) of molecules to their biological activities through statistical analyses. This technique is helpful in determining what properties and features of a molecule are responsible for the experimentally observed biological activity.⁹² Molecular descriptors are numerical values that can be employed for characterizing the properties of the molecules.

$$\emptyset = f(x) + error$$

The biological activity (\emptyset) is defined as a function of x , where x are the molecular descriptor values. Since its conceptual implementation by Hansch in 1969, the field of QSAR has witnessed significant changes and development.⁹³ The applications of QSAR methodology extends from medicinal chemistry to toxicology and risk assessment.⁹⁴

There are two kinds of QSAR models that can be constructed depending upon the type of the predictor attribute: continuous or categorical.⁹⁵ Continuous QSAR models are built to predict

continuous numerical properties such as IC_{50} , K_D values. In contrast, categorical QSAR models are used for classifying instances into distinct classes using classification algorithms. A general purpose QSAR workflow has been shown in Figure 1.4 listing the crucial model building steps.

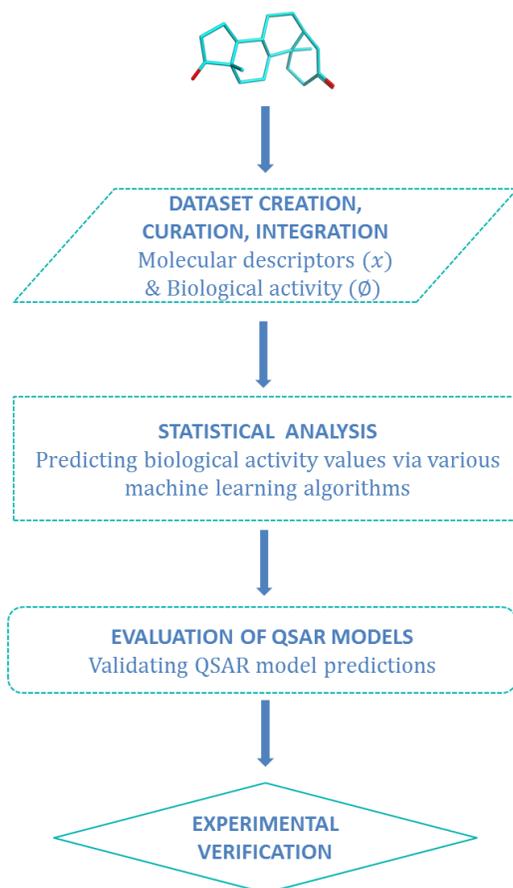


Figure 1.4: A general QSAR workflow, the dataset is generated followed by statistical analyses and evaluation followed by experimental verification to predict activity/inactivity of a molecule

1.2.2.2 Evaluating QSAR models

The evaluation of the models is an integral and essential step in QSAR analysis.⁹⁶ The predictions generated by QSAR models need to be validated for their statistical accuracy. This can be done in either or all of the following ways:

1. *Internal validation*- Techniques such as cross-validation can be employed to evaluate the robustness of the model. This can be extended into several validation folds such as n-fold validation, which is very powerful in determining statistical accuracy and model correlation.⁹⁷
2. *External validation*- For external validation, a part of the dataset is separated out and is not used in the model training process. This external test set is then supplicated to assess the model performance.⁹⁷
3. *Blind external validation*- QSAR models generate predictions for an independent dataset, which is then used for model quality and performance assessment.⁹⁸
4. *Data scrambling*- This technique is used to determine any correlation between the response variables and features, moreover to ascertain any probability of observing a correlation. With the response variable randomized, the QSAR model should now perform poorly, since no meaning can be inferred from such models. Thus, this technique can detect and quantify correlations between the response variables and features.⁹⁹
5. *Estimation of the applicability domain*- The uniformity in distribution of molecular descriptor values can be evaluated between both the training and test set instances by applicability domain estimation. Utilizing this information, the exactness of predictions can be estimated for any random dataset with minimal similarity consideration used for model validation. By the use of a leverage-based method (such as Euclidean distance), those test set instances, in comparison to the training set instances either exceeds or are below the predefined limit can be classified to be 'out of the applicability domain'. The accuracy of predictions made for such instances cannot be considered reliable.¹⁰⁰

1.3 Cheminformatics

Cheminformatics can be defined as an amalgamation of computational and informational techniques used for the storage, retrieval and mining of chemical information. Cheminformatics complements CADD, by providing statistical support to the decision making process of chemical compound screening and selection through the mining of chemical information to obtain statistical trends that can be correlated with experimental observations.¹⁰¹⁻¹⁰² Cheminformatics can be described as a two part process: first being *encoding* or representation of molecular structure by vector of features. The latter process is termed as *mapping* or empirically relating the features to a property of interest such as physicochemical, bioactivity, ADMET. Machine learning algorithms are employed for the task of mapping.¹⁰³

1.3.1 Machine learning in Cheminformatics and Classification algorithms used

Machine learning represents a conventional data analysis technique that automates computational model building that can be used for instance-based learning from available data and finding statistical trends.¹⁰⁴ For the mapping of dependent variables to the features, both supervised and unsupervised machine learning algorithms are used.

A composite of both unsupervised and supervised machine learning algorithms were employed for the construction of QSAR models. A total of 7 algorithms were used in the development of QSAR models to model the adverse drug responses of clinically relevant AR mutants and to predict new ones. These were: DecisionStump, OneR, RandomForest, Bagging, Dagging, IBk, and LibSVM of the WEKA datamining software.¹⁰⁵⁻¹¹⁰ Most of these algorithms have been widely used in the development of QSAR models and their merit over other machine algorithms lie in the low training error, simplicity and better interpretability. The algorithms used for QSAR model development have been described in the subsequent text.

1.3.1.1 DecisionStump

DecisionStump comprises of one-level decision tree. It is a weak learner with a very simple structure built up of one single attribute (as shown in figure) split that can be combined with other learning algorithms for a better accuracy.¹¹¹ In an example shown below, the decision tree has been created on the basis of an attribute value of $a \leq 50$, if this condition is satisfied the instance would be classified into class X, else class Y. (See Figure 1.5)

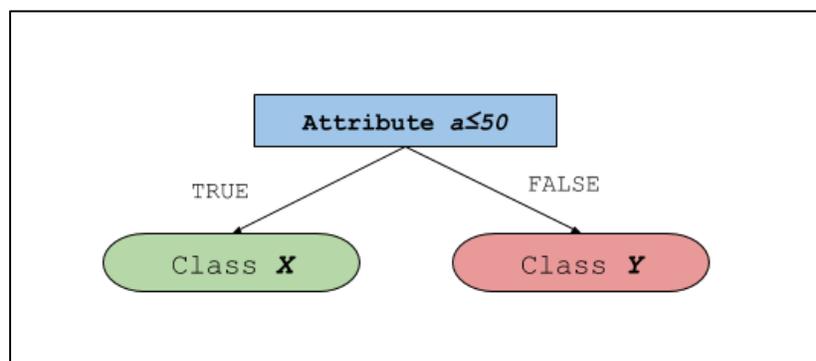


Figure 1.5: Workflow of DecisionStump algorithm, classifying based on a 1-level decision tree

1.3.1.2 OneR

OneR or one rule is a machine learning algorithm that generates one rule for a predictor value present in the dataset and then selects the rule with the minimum error, using it as the 'rule'.¹¹² A frequency table is generated for each of the predictors, along with the error rates. In Figure 1.6, an example of two such rules has been shown. According to *Rule 1*, for the attribute a , if the value is ≥ 50 , how accurately would it classify the instances is determined by error frequency table. Clearly, the error rates in predicting using *Rule 1* are lower as compared to *Rule 2*, therefore it would be selected for building up the classifier.

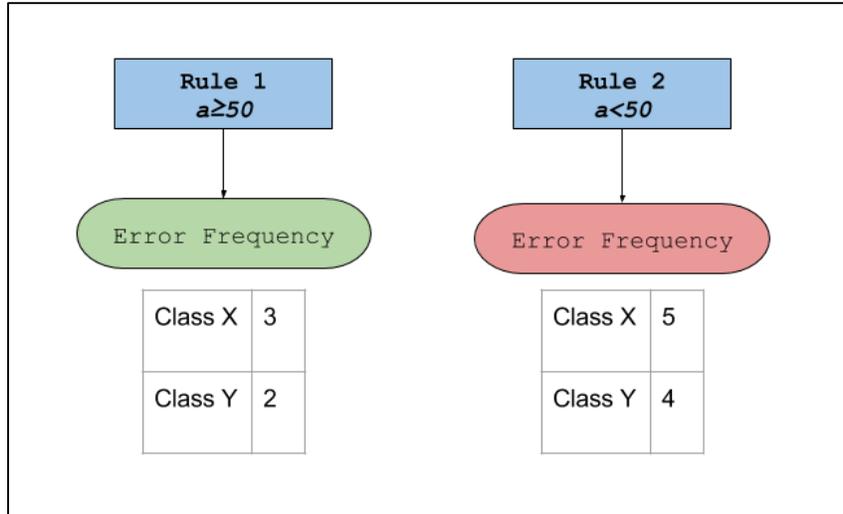


Figure 1.6: One-rule based classifier algorithm, selects the rule for classification with minimum error frequency

1.3.1.3 RandomForest

The RandomForest algorithm as the name suggests is an algorithm that operates upon a collection of decision trees. It is an ensemble machine learning method that generates predictions followed by subsequent majority voting. This involves supervised learning, by mapping of the input variables into discrete categories.¹⁰⁶ (Figure 1.7)

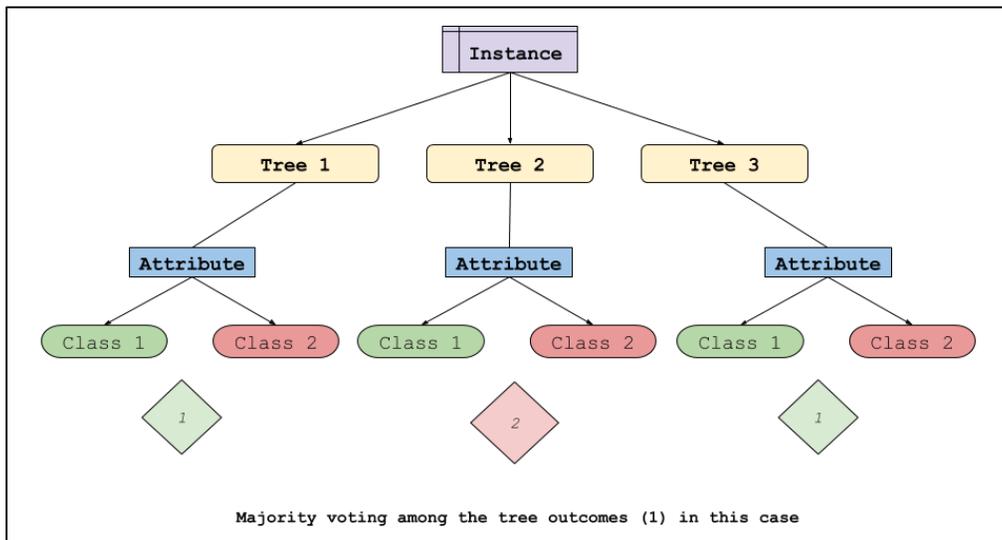


Figure 1.7: RandomForest algorithm implementation

1.3.1.4 Bagging

Bagging or bootstrap aggregation is an ensemble machine learning approach that aims at improving the prediction accuracy by combining multiple classifiers. For example, a dataset with n instances is divided into subsets drawn with m replacements. These sets are supplied to the classifiers C , to generate predictions which are further filtered through majority voting.¹¹⁰ (See Figure 1.8)

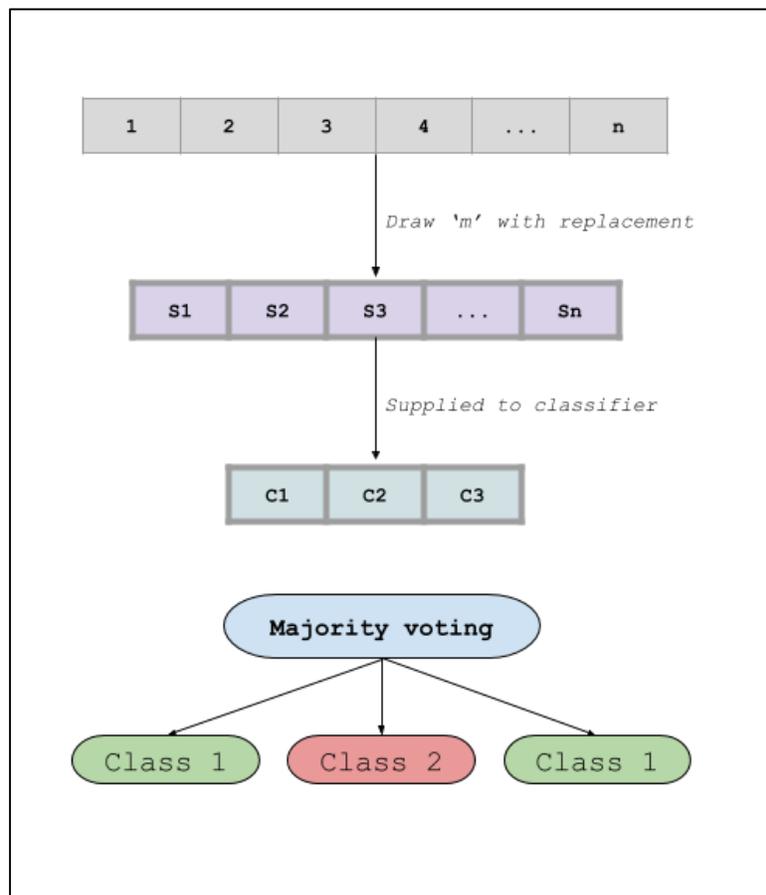


Figure 1.8: Classification using Bagging algorithm workflow, splits the dataset into smaller subsets which is supplied to classifier followed by majority voting

1.3.1.5 Dagging

Dagging bears a degree of similarity to the bagging algorithm. The data is divided into several folds and supplied to the base classifier which generates the prediction values for each of the instances. The final outcome for each instance is determined through majority voting.¹¹⁰ (Figure 1.9)

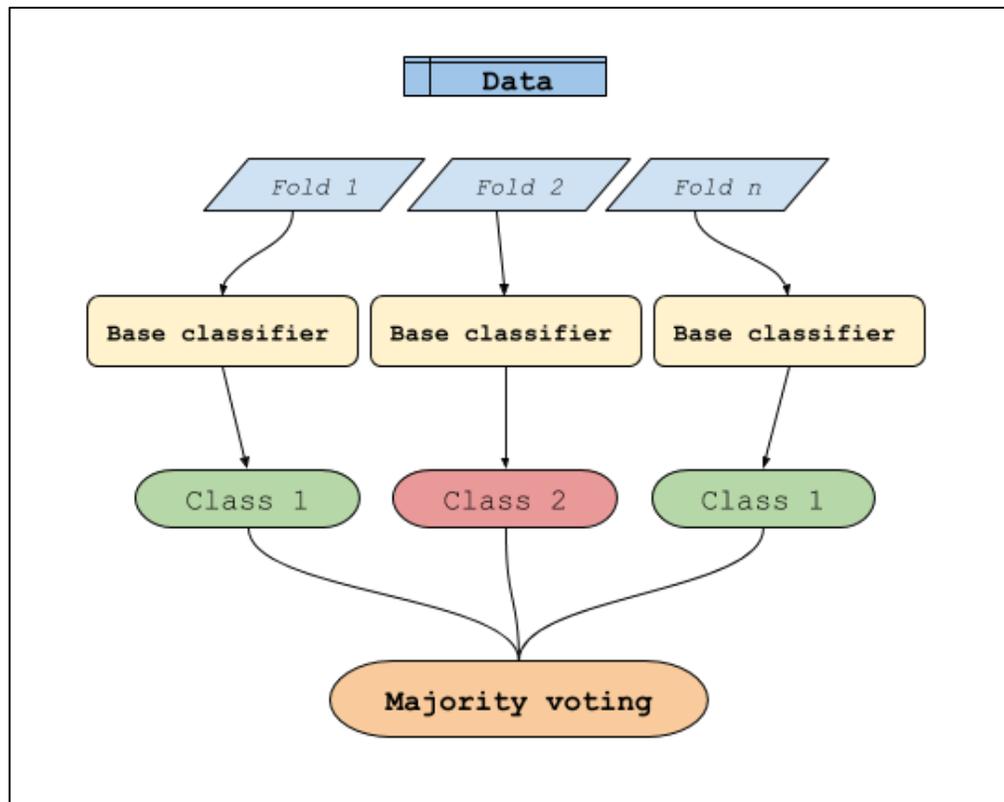


Figure 1.9: Dagging algorithm workflow

1.3.1.6 IBk

IBk or 'k' nearest neighbor classification algorithm is a non-parametric approach that looks up the 'nearest' neighbors of a test set instance depending upon a majority vote of the training set instances.¹¹³ Distance functions such as Euclidean or Manhattan distances are generally used to

measure the distance between the test instance and the training set instances to determine the nearest neighbors and the class of nearest lying training set instances is assigned to the test set instance.¹¹⁴ The class labels are assigned to the test instance based upon the threshold (number) of nearest neighbors defined. (See Figure 1.10) As shown in the figure for a value of $k=3$, the training space is searched to find at least 3 neighbors that lie close to the test instance for classifying the class of the test case.

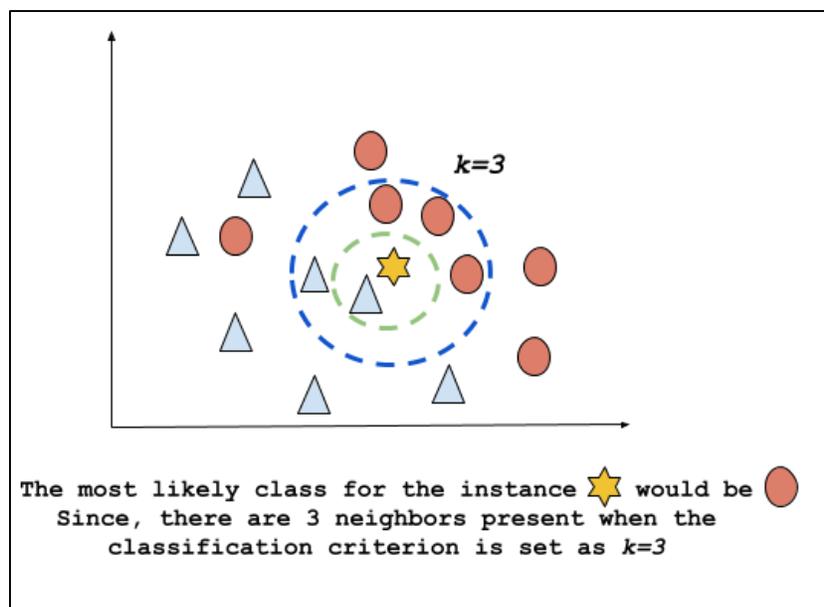


Figure 1.10: Classification using k nearest approach (IBk), shown is the classification of a test set instance when $k=3$

1.3.1.7 LibSVM

LibSVM algorithm implemented through WEKA is based on support vector machines (SVM) algorithm used for classification and regression purposes.¹⁰⁸ The classification is based on a small fraction of training instances called support vectors, which can be used to discriminate between the two categories. The method is very sensitive in determining outliers and exhibits

resistance to overfitting. A high value of the hyperplane that provides maximum separation between two support vectors signifies a well separated system of distinct categories. (Figure 1.11) SVM can be employed in both linear and non-linear classification problems by using kernel trick to map dataset instances into high-dimensional feature spaces and achieve maximum separation between support vectors.¹¹⁵⁻¹¹⁶

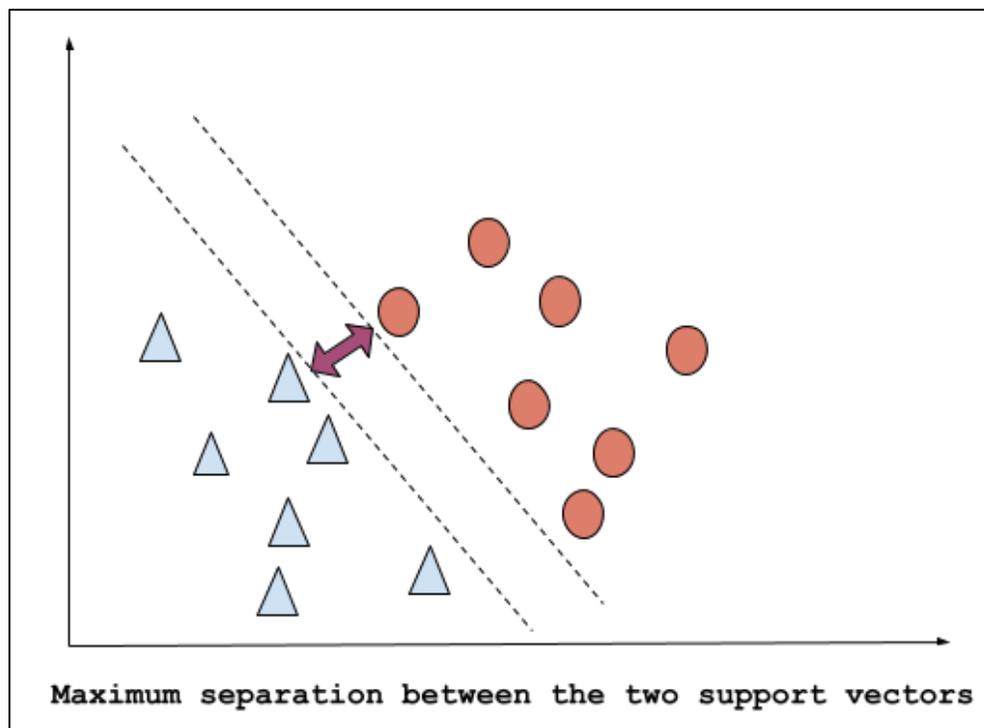


Figure 1.11: LibSVM algorithm workflow for classification

1.3.2 Exploratory data analysis for attribute selection

Machine learning algorithms are also used for exploratory data analysis. Exploratory data analysis provides valuable insights about the data. Particularly in computational chemistry, QSAR models are built upon molecular descriptors that predict biological outcomes. A large number of molecular descriptors can be quantified and most of the computed values may

possibly be highly correlated.¹¹⁷ The removal of highly correlated, redundant, and statistically insignificant attributes enhances prediction efficiency and decreases the computational complexity.

A wide range of attribute or feature selection methods are available that can be used depending on the nature of the analyses. For this study, attribute prioritization was a critical step to obtain statistical inferences about the dataset and has been further described in Section 2.3.3. To determine the important attributes, Boruta package in R programming language was used.¹¹⁸ It depends on random forest classification algorithm which gives an intrinsic measure of each attribute, a Z-score that can be used to compare the importance of different attributes in the dataset. The caret (classification and regression training) package was employed for feature evaluation. Recursive feature elimination implemented to determine the optimum number of features required for model building, with the lowest error rate.¹¹⁹

Chapter 2: Cheminformatics platform development

An *agonist* can be defined as a substance that specifically binds to a receptor with high affinity and elicits the biological response, whereas, an *antagonist* binds to a receptor but inhibits the biological response.¹²⁰ In prostate cancer, androgen deprivation therapy (ADT) is administered to patients through anti-androgens which act as antagonists. This antagonist response is characterized by reduction in PSA levels. However, in cases where AR becomes promiscuous, the response of anti-androgen is modified into *agonist*.¹²¹⁻¹²² This conversion fails to elicit the anticipated response antagonism, rather increasing levels of PSA are observed.¹²³ Figure 2.1 illustrates the initial cheminformatics pipeline that was developed for the purpose of this study. The first phase of the workflow included quantification of protein-ligand interactions captured through the molecular docking with known anti-androgens as well as the native AR ligand-DHT. This quantification was achieved through Glide-per-residue scores as well as the novel 4-D Inductive descriptors developed in-house. The experimental characterization of previously reported AR mutants, which were detected via cell-free DNA (cfDNA) sequencing unveiled new gain-of-function scenarios.¹²⁴ The therapeutic response was therefore, categorized into agonist and antagonist classes. These response classes were then assigned to the various protein-ligand complexes, via a nominal attribute of '*activity*'. Upon activity assignment, the prepared dataset was statistically analyzed through machine learning to construct QSAR models, therefore predicting the therapeutic responses. However, this pipeline had certain short-comings, which have been shown in red in Figure 2.1 and listed below:

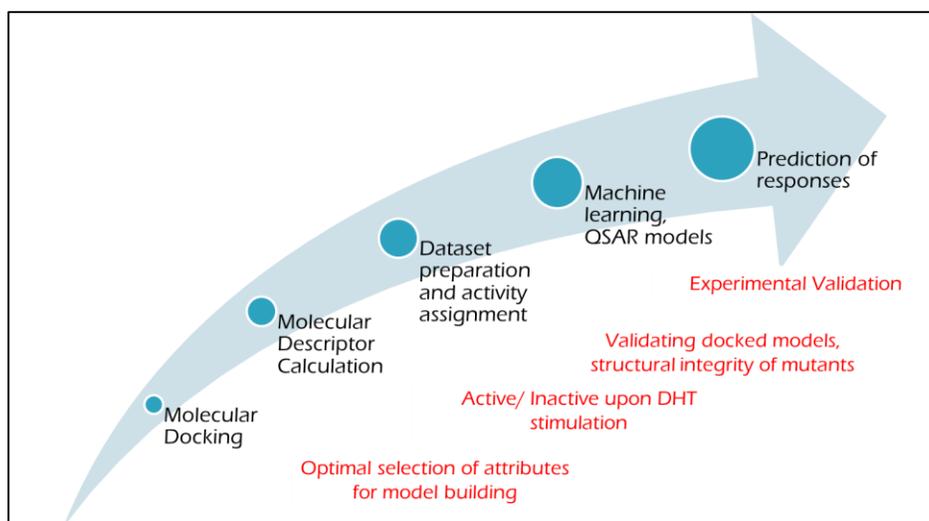


Figure 2.1: Initial pipeline used for predicted mutant - anti-androgen responses

1. The initial pipeline lacked an attribute filter that is necessary for the optimal attribute selection resulting in high accuracy of the predictive models. This aspect has been further described in Sections 2.3.3 and 2.3.4.

2. Only a single QSAR modeling layer existed earlier, accountable of generating therapeutic response predictions of agonist/antagonist. An additional stratum was added, that predicts whether a given mutant will be active or inactive upon DHT stimulation.

3. The protein-ligand complexes were produced as a result of molecular docking experiments. However, the structural viability was unknown which required verification and MD simulations were used to assess the system equilibrium and stability.

4. The cross-validation with experimental characterization results was essential, to assess the predictive power of the QSAR models generated for mutant – anti-androgen responses.

Therefore, the final pipeline evolved that integrates all the aforementioned aspects and has been shown through a flowchart in Figure 2.2.⁹

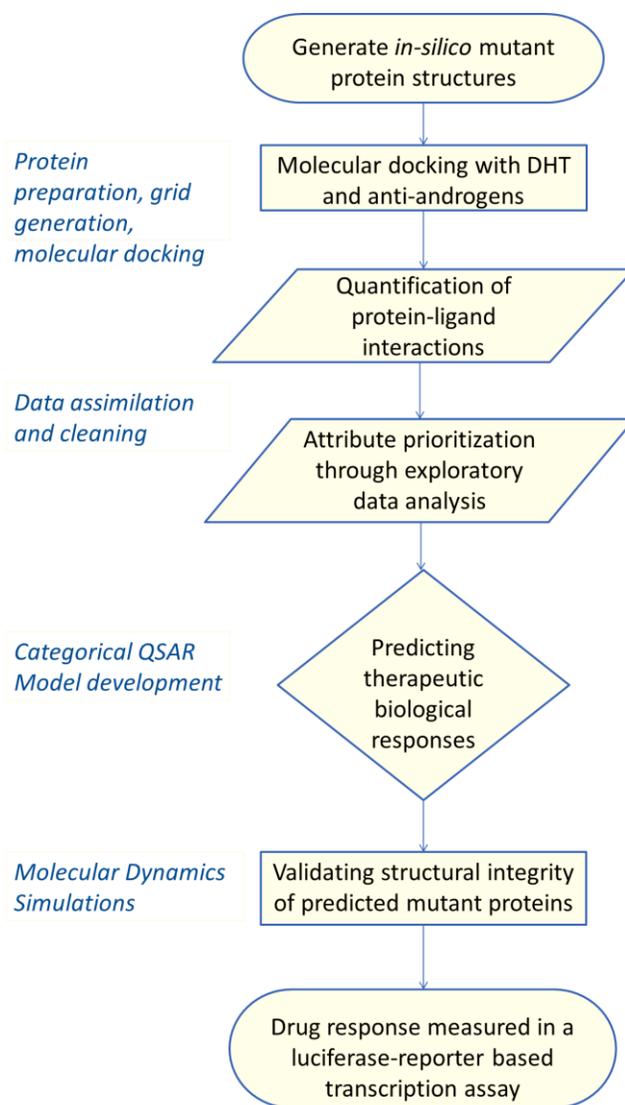


Figure 2.2: Final functional pipeline developed for studying mutant- anti-androgen responses

The first section of the pipeline represents computational modeling of the proteins, primarily through generation of *in-silico* protein structures. This was followed by molecular docking of ligand molecules into the protein models for capturing the binding interactions of various mutants and ligand molecules. The quantification of these interactions by molecular descriptor computation was the data aggregation operation. Exploratory data analysis was then employed

for data assimilation and cleaning. With the aid of machine learning algorithms, various categorical QSAR models have been built, followed by cross-validation of the predictions of the experimental outcomes. To ensure integrity of the predicted mutant structures of the AR, as well as validating docked poses, the corresponding protein-ligand complexes were subjected to molecular dynamics (MD) simulations. The final block of the pipeline (Figure 2.2) represents the experimental verification of predicted therapeutic responses through in-vitro experiments.

2.1 Datasets

A total of 24 mutants located across the ligand binding domain (LBD) of the AR were recently characterized by our group. Upon detailed investigation of the anti-androgen treatment concentrations, statistical trends were used to classify their overall response as either agonists or antagonists. We built the wild-type AR structure based on PDB ID: 1Z95 crystallographic structure by adding missing hydrogen atoms and by performing energy minimization.⁴⁰ The resulting energy equilibrated wild-type structure was used as a template to generate mutant protein structures. The initial structure setup for wild-type and mutants were carried out with Molecular Operating Environment (MOE) 2015.1001, that is a computational chemistry software package.¹²⁵

2.1.1 Training dataset

The training set was built comprising of the 24 AR mutants and the wild-type AR earlier studied by our group, since their response trends were conclusive classifying them as agonists or antagonists. (See Table 2.1 and Appendix A for more details) The QSAR training set was constructed containing 84 mutant– anti-androgen complexes along with the biological activity classes of either agonist or antagonist. The amino acid residue within the wild-type sequence was modified corresponding to the mutated residue of the AR, via Residue scan module of MOE.¹²⁶

To address any conformational issues as result of the substitutions, energy minimization was performed with Amber10 force field (set of molecular mechanical forcefields to simulate the biomolecules).¹²⁷

<i>S. No.</i>	<i>Wild-Type Residue</i>	<i>Residue #</i>	<i>Mutated Residue</i>
1	L	702	H
2	V	716	M
3	V	731	M
4	W	742	L
5	W	742	C
6	H	875	Y
7	H	875	Q
8	F	877	L
9	T	878	A
10	T	878	S
11	D	880	E
12	L	882	I
13	S	888	G
14	D	891	H
15	E	894	K
16	M	896	V
17	M	896	T
18	E	898	G
19	T	919	S
20	T,S	878,889	A,G
21	T,D	878,891	A,H
22	H,T	875,878	Y,A
23	F,T	877,878	L,A
24	H,T	875,919	Q,S
25	WILD-TYPE	-	-

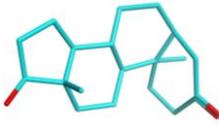
Table 2.1: List of clinically reported AR wild-type substitutions

2.1.2 Test dataset

A total of 28 amino acid residues located within the LBD were substituted by 19 natural amino acid residues. A parallel approach such as the one that was employed in the training set was used to generate 532 protein models. This approach works well for generating single point mutants. To generate double point mutants, i.e. with a mutation site limit of 2, over $1.00e+007$ structures would be generated. The creation of double mutants is exceedingly computationally expensive, given the number of possible permutations. The Residue Scan module allows a maximum site limit of 6, which would generate about 64 million mutant structures.

2.1.3 Ligand molecules used

We used a ligand set containing the native ligand DHT, currently used anti-androgens as well as other experimental anti-androgens, which are subjects of ongoing clinical trials. The structure of ODM-201 has not been shown. Table 2.2 lists the ligand molecules used:

<i>S. No.</i>	<i>Molecule</i>	<i>Structure</i>
1	DHT	
2	ARN-509	

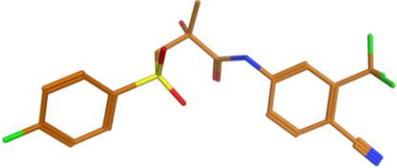
<i>S. No.</i>	<i>Molecule</i>	<i>Structure</i>
3	Bicalutamide	
4	Enzalutamide	
5	Hydroxyflutamide	

Table 2.2: Ligand molecules used for docking, containing native ligand DHT and other anti-androgens

2.2 Molecular docking protocol

The energy equilibrated wild-type structure generated based on PDB ID: 1Z95 together with the mutant structures were docked with the ligand molecules shown in Table 2.2.

2.2.1 Protein structure preparation

The protein structures were prepared with the Protein Preparation Wizard within Maestro suite version 10.1.013, Release 2015-1 (Schrödinger, LLC). All solvent molecules were removed, followed by hydrogen bond assignment, adjusting the bond order. A restrained energy minimization was carried out using the OPLS_2005 force field.¹²⁸ Subsequently, 22 of the

structure models were eliminated during the protein preparation, given certain substitutions cause steric clashes and structural disruption of proteins.

2.2.2 Receptor grid generation

Receptor grids were generated corresponding to the active site within the protein structure, with a reduced Van der Waals radius scaling factor to 0.8 in order to soften the potential for non-polar parts of the receptor. The inner box of the grid defines the volume that the ligand center explores during the exhaustive site-point search across the X, Y and Z coordinates which was defined as 20 Å.

2.2.3 Glide XP docking mode

Glide is a ligand docking program that was used for docking the anti-androgens to the ABS of AR. There are 3 different docking modes available: HTS (High Throughput Screening) that scans large (>1 million molecules) compound libraries; SP (Standard Precision) that docks large number of compounds with enhanced accuracy. As compared to the SP (standard precision) mode, the XP (extra precision) mode performs more extensive sampling through a more sophisticated scoring function, with higher requirements for ligand-receptor shape complementarity.⁷⁶⁻⁷⁷ This helps in the reduction of false positives that the SP mode would not have otherwise penalized.¹²⁹ The Glide XP program was used to dock the compounds into the ABS of the AR structures implemented in the Maestro suite.

2.3 Molecular descriptor computation

Molecular descriptors have been used to capture chemical features of the molecules, over various dimensions of structural representation.¹³⁰ These descriptors range from 0-D to 6-D:

<i>Category</i>	<i>Description</i>
0-D	Count descriptors, provides no information about the molecular structure. Example: atom and bond counts.
1-D	Structural information. Example: SMILES
2-D	Topological descriptors, used to describe the topology of the molecule. Example: connectivity indices, 2D fingerprints
3-D	Based on the 3-D structures of the molecules (Cartesian or internal coordinates). Example: dihedral angles, bond angles
4-D	Based on trajectory data of a molecular dynamics simulation. Example: Volsurf, GRID derived descriptors
5-D	Based on induced-fit parameters, representation of various induced-fit models. Example: induced fit docking scores/parameters
6-D	Based on information obtained from solvation states along with 5-D descriptors. Example: Quasar

Table 2.3: Different categories of molecular descriptors used in QSAR modeling

Molecular descriptors have been widely applied to diverse QSAR modeling pipelines.¹³¹ The computation of molecular descriptors was carried out in Glide (Schrödinger) and MOE (Chemical Computing Group).^{76-77, 125}

2.3.1 Glide per-residue energy scores

Glide per-residue energy scores were calculated for residues lying within 10 Å of the ligand molecule. 4 different parameters calculated for all those residues within the 10 Å region from the ligand, which were:

1. *vdw*: This residue-ligand interaction parameter, scored by measuring the Van der Waals energy of interaction between the pair.
2. *coul*: This parameter measures the Coulomb's energy of interaction between the pair.
3. *eint*: It represents the total non-bonded interaction energy, represented as a sum of *vdw* and *coul* energy parameters.
4. *hbond*: The H-bond per-residue interaction parameter is the total sum of individual H-bond scores between the ligand and a single residue. This is influenced by atom-types and geometries involved in H-bonding.

2.3.2 4D-Inductive descriptors

For our study, the quantification of protein-ligand interaction was of prime importance: given, substitution of wild-type AR residues by a single amino acid could result in a complete switch of biological responses. This quantification was not possible through the available small molecule descriptors. In order to quantify interactions between proteins and ligands, novel 4D-inductive descriptors were developed based on previous 3-D models of inductive descriptors. The developed descriptors account for the protein structure as well.

On the basis of the 3D models of inductive (polar) effects (σ), steric effects (R_s , Abs_R_s), 'inductive' electronegativity (χ), 'inductive' charge (Q) and molecular capacitance, a range of novel 4-dimensional (4D) inductive descriptors has been developed for this study to quantify receptor-ligand interactions at atomic level.¹³²⁻¹³⁴ The theory and mathematical formalism of the 'inductive' descriptors have been widely applied to various studies such as estimation of hemolytic C-H bond dissociation enthalpies etc.¹³⁵⁻¹³⁷ All the 'inductive' descriptors possess physical meaning and are calculated based on atomic properties like covalent radii, electronegativity and interatomic distances. This enables their use in QSAR analyses of huge

chemical compound datasets and also, for larger molecular systems such as proteins and complexes.¹³⁷⁻¹³⁹ The 4D inductive descriptors were implemented in the Scientific Vector Language (SVL)¹⁴⁰ and computed by our in-house script integrated in MOE. The general formulae used to compute the inductive descriptors have been listed below.¹³⁶

$$\sigma^* = \sum_{i=1}^n \frac{\Delta X_i R_i^2}{r_i^2}$$

Inductive constant of any substituent at reaction center

$$R_S = \sum_{i=1}^n \frac{R_i^2}{r_i^2} = \sum_i R_{S_i}$$

Steric parameter computed for the i-th atom into overall R_S group value

$$q_i = \sum_{j, j \neq i}^{N-1} \frac{\Delta X_{j-i} (R_j^2 + R_i^2)}{r_{j-i}^2}$$

Pair inductive charges at an atom 'i' computed through pair inductive interactions with other 'j' atoms of the molecule

The developed 4-D Inductive descriptors have been further described in Section 3.2.1.⁹

Three categories of descriptors were developed and calculated for all receptor-ligand complexes: (1) $R_S_L_R$ (steric effect R_S), $Abs_R_S_L_R$ (absolute steric effect Abs_R_S), σ_L_R (inductive effect σ), and Q_L_R (overall charge Q) descriptors were calculated to measure, the steric, absolute steric, inductive electronegativity and overall charges respectively, the cumulative influence of all ligand (L) atoms on all receptor (R) atoms within 10 Å receptor region surrounding the ligand;

(2) From the overall receptor (R) and ligand (L) range, specific amino acid interactions were quantified using the AA# (amino acid residue number) descriptors. AA#_Rs_AA_R, AA#_Abs_Rs_AA_R, and AA#_σ_AA_R descriptors were computed to quantify the steric and inductive effects of all atoms of mutated amino acid residues (AA#) relative to the wild-type; and

(3) The hybridized states of the receptor (R) and ligand (L) molecules were quantified by R_Ah_L_Ah descriptors, as the inverse square of interatomic distances, the interactions between all receptor atoms within the 10 Å cut-off region in all possible hybridized states (R_Ah) and all ligand atoms in their hybridized states (L_Ah).

The significance and interpretation of our novel 4D inductive descriptors in discriminating the behavior of AR mutants relative to the AR wild-type are highlighted in Section 3.2.

2.3.3 Attribute pruning and prioritization

Attribute pruning and selection is one of the most important steps of QSAR modeling that removes highly correlated attributes, filters out outliers and other ambiguities in the dataset such as missing, or redundant points. The exploratory data analysis workflow was created in R programming language and implemented in the RStudio environment.¹⁴¹⁻¹⁴² The Boruta algorithm was used for attribute prioritization by statistical importance ranking.¹¹⁸ The algorithm establishes relevance by comparing the importance of actual attributes against random probes (shadow attributes). It uses a wrapper built around the random forest classification algorithm implemented in the R randomForest package.¹⁴³ The random forest algorithm is considerably quick and easy to implement that gives a numerical estimate of attribute importance. Being an ensemble method, it encapsulates numerous unbiased decision trees and their consensus voting for deducing the classification. The Boruta algorithm encompasses the following steps:

1. Dataset expansion, by addition of shadow attributes followed by shuffling of added attributes alienates their correlation with the response variable.
2. Accumulation of Z scores computed in the random forest classifier run.
3. Determination of maximum Z score among the shadow attributes (MZSA); those hits which score better than the MZSA are now assigned as IMPORTANT attributes.
4. Attributes whose significance was undetermined land up in the UNDECIDED category.
5. Attributes with lower significance than MZSA were assigned to the UNIMPORTANT category, which must be removed from the dataset.
6. Attributes with higher significance than MZSA were assigned to the IMPORTANT category.
7. All shadow attributes were removed.
8. Procedure was repeated until the importance was assigned into only two categories: IMPORTANT and UNIMPORTANT or the algorithm reached the maximum limit of random forest runs, previously assigned.¹¹⁸

2.3.4 Attribute selection

Recent datasets have been described with way too many variables for model building. A sufficient attribute pool is the prime prerequisite before operating on the dataset with machine learning algorithms. The caret (Classification And Regression Training) package developed in R programming language was incorporated into our exploratory data analysis pipeline.¹¹⁹ Wrapper method such as recursive feature elimination (RFE) was applied to evaluate multiple models, involving the addition or removal of attributes to find the optimal combination which would maximize the performance of the model. RFE serves 3 tasks: attribute selection, model fitting and performance evaluation.¹⁴⁴ Furthermore, through resampling of the attributes, the performance estimates were obtained that reflects the variation due to attribute selection.

2.4 QSAR model development

Upon completion of the attribute pruning, prioritization and selection, the dataset was assimilated and prepared. The key factor defining the type of QSAR models to be developed was the nominal attribute of ‘activity’ which held two class responses of either AGONIST (+1) or ANTAGONIST (-1). These models were built upon the algorithms described in Section 1.3.1.

2.4.1 Model building and consensus vote approach

The KnowledgeFlow application in the WEKA suite, was used to design the procedural flow for generating QSAR models, as shown in Figure 2.3 below.¹⁰⁸ Machine learning algorithms were implemented for QSAR model building, including Bagging, Dagging, Local-lazy method (IBk), DecisionStump, LibSVM, OneR, and RandomForest. These algorithms generate binary predictions for a nominal attribute of ‘activity’. The binary predictions were either +1 (Agonist) or -1 (Antagonist) that were produced for all the instances.

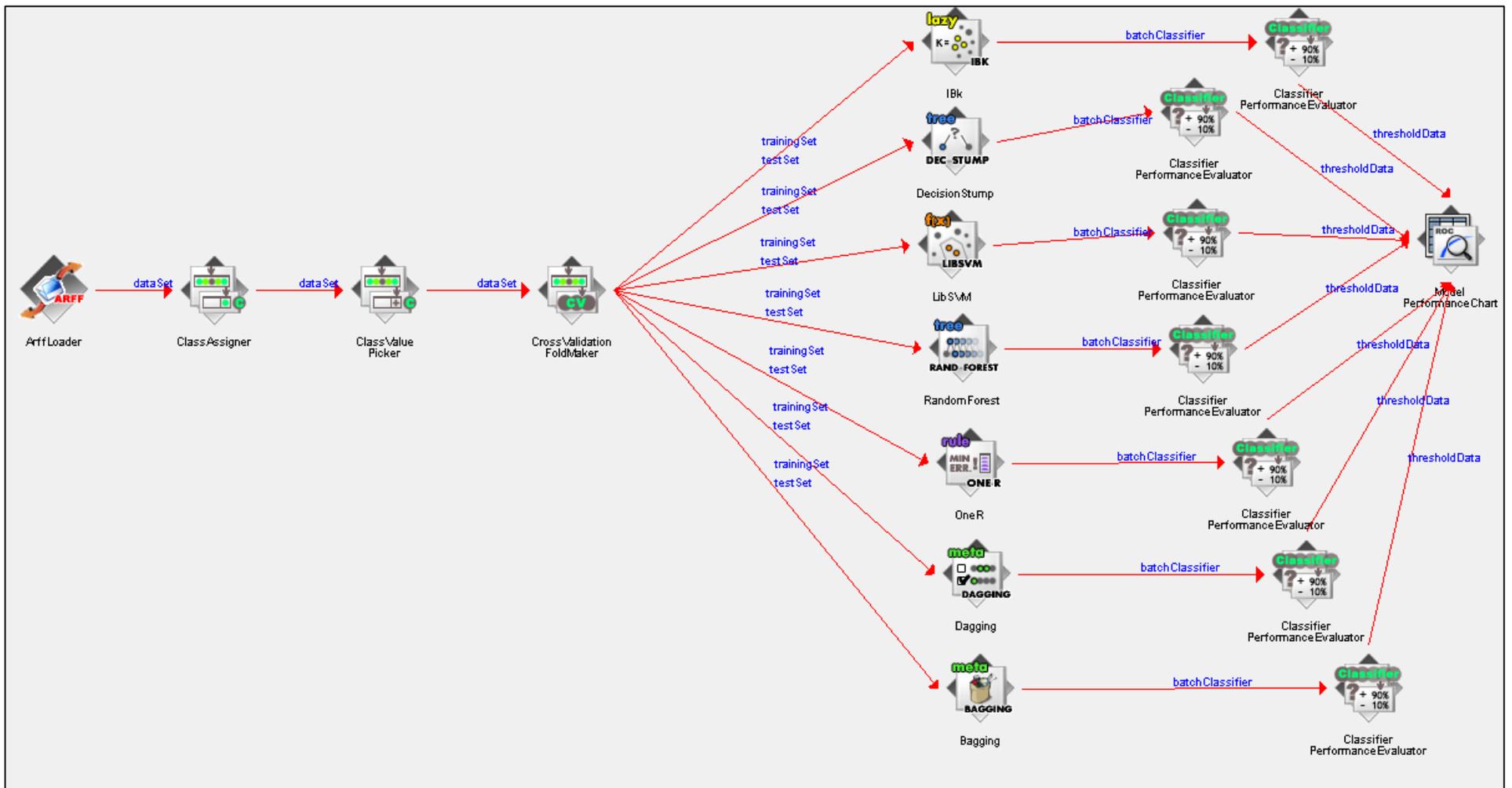


Figure 2.3: WEKA knowledge workflow implemented

Consensus voting approach is applied to the various predictive models generated, that gathers all solutions and employs majority voting to determine the predicted category of an instance. This approach maximizes the prediction performance and also describes the use of only one single algorithm for predicting may not be the ideal case, which may produce biased predictions.

2.4.2 Evaluation of QSAR models

A 10-fold cross validation approach was used for evaluating the QSAR model. The dataset was divided into k equal subsets, then $k-1$ subsets are treated as training set whereas the remaining 1 subset is retained for testing.¹⁴⁵ Since, a 10-fold cross validation was to be performed the process was repeated 10 times, with each of the k subsets being used as a test set. The results of these folds (number of times the test was repeated) can be averaged or aggregated yielding a single estimation.¹⁴⁶ Furthermore statistical measures of the model performance reflect the accuracy and efficiency for a binary categorical classification testing problem, like ours. The following parameters were computed:

1. *Sensitivity*: Sensitivity is the proportion of correctly identified true positives (TP) to the probability of a positive outcome for the test. TP in this case is the number of correctly identified +1 (agonist) responses.¹⁴⁷

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

2. *Specificity*: Specificity is the proportion of correctly identified true negatives to the probability of a negative outcome for the test. TN in this case is the number of correctly identified -1 (antagonist) responses.¹⁴⁷

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

3. *Precision*: Precision or positive predictive value (PPV) is the ratio between the number of true positives and the total number of positive outcomes (both true and probable).¹⁴⁷

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

4. *Accuracy*: The proportion of the correctly identified positive and negative outcomes (true positives and negatives), to all the outcomes of the diagnostic test, is the statistical measure of accuracy.¹⁴⁷

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

5. *ROC space*: The accuracy of a testing problem such as categorical classification largely depends on how well the true and false positives are distinguishable. Thus, the Receiver Operator Characteristic (ROC) curve was generated by plotting the values of the sensitivity against 1-specificity (false positive rate, FPR).

$$FPR = \frac{FP}{FP + TN}$$

A perfect ROC curve would cover an area of 100 % while area below 50% coverage denotes random outcomes which are statistically insignificant.¹⁴⁸

2.4.3 External test set validation

The predictions were generated for two external datasets, one dataset comprising of the in-silico engineered mutants and the second set containing known AR mutants complexed with the experimental anti-androgen, ODM-201. The mutant-drug responses were unknown at the time of model construction and testing for ODM-201 dataset.

With the absence of the experimental anti-androgen ODM-201 in the QSAR training dataset, the external validation could examine the applicability and adaptability of the models. Predictions were generated, using similar approach as for the in-silico mutant dataset.

2.5 Molecular dynamics simulations analyses

To evaluate the structural integrity and system equilibrium of the mutant protein-drug complex, molecular dynamics (MD) simulations were carried out with Amber and Gaussian09 packages via computing facilities provided by Westgrid.¹⁴⁹⁻¹⁵⁰

2.5.1 Steps involved in MD simulations

The protein-ligand complex was checked again for any issues. The macromolecular complex was assigned the protonation state through Protonate3D application of MOE, and AMBER10 (Assisted Model Building with Energy Refinement) force field was used.¹⁵¹⁻¹⁵³ The ligand molecule was extracted separately from the protein-ligand complex and is saved as a PDB file. This file was then converted into XYZ format, using VMD; that specifies the ligand geometry represented by atom number aligned with the Cartesian coordinates.¹⁵⁴ The different steps of MD simulations have been described in subsequent sections.

2.5.1.1 Ligand geometry optimization and atomic charge assignment

The geometries of the ligand molecules were fully optimized using the Gaussian09 suite of programs, accessed through Westgrid's Grex cluster. The Gaussian electrostatic potential (GESP) file was generated that bear electrostatic potential information around the ligand molecule, computed quantum-chemically using HF/6-31G optimization (basis set to add flexibility and polarization functions to atoms).¹⁵⁵ The job was submitted on the Grex cluster with a wall time of 168 hours on 2 processors with a 2000 MB memory allocation.

The atomic charges were assigned through Amber's Antechamber program based on restrained electrostatic potential (RESP) fitting and generalized amber force field (GAFF) atom types to the ligand molecule.¹⁵⁶⁻¹⁵⁸ The hydrogens were then stripped off the protein, used as an input for the TLEAP program. All subsequent processes including Antechamber's atomic charge assignment were carried out on the Jasper cluster, which is intended for serial and MPI-based parallel computing.

2.5.1.2 Protein-ligand forcefield assignment and energy minimization

TLEAP program reads the pre-determined coordinate files (PDB) and generates a topology file. This program assigns the protein-ligand force fields and neutralizes the system by adding counter ions either Na⁺ or Cl⁻ depending on the total charge of the system. Then, the system was solvated in a TIP3P 10.0 Å water box. Energy minimization was performed for removal of any steric clashes by gradually decreasing the restrain weights from 50 to 0 kcal/mol-Å². The minimization is carried out with stepwise decrement of restrain weights since a sudden change, may disrupt the structural integrity.

2.5.1.3 Heat, pressure application and production run

The system was heated from 100 K to 300 K with a fixed protein structure, and restraint weight of 10.0. After heating the system up to 300 K, pressure measuring 1 ATM was applied to the system reducing the restraints from 10.0 to 0 in a 3-step process.

The 25 ns production run was initiated on 24 processors across 2 nodes on Westgrid's Jasper cluster with a wall time of 72 hours and 2000 MB memory allocation. On an average, the production run completed in 40 hours.

2.5.2 RMSD and contact frequency analysis

Root mean square deviation (RMSD) of atomic positions was evaluated between the initial and final points of MD simulations. This measure of average distance between the backbone atoms of superposed protein structures, gives an insight about the overall structural stability and integrity of the structure. The conformations of the mutants were oriented in space in order to optimally superimpose the backbones upon each other. This molecular fitting measured by the distance of atoms between the two superimposed protein structures, and can be described by the following formula:

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

RMSD computation between two receptor-ligand complexes

where, N_{atoms} : number of atoms over which RMSD is measured and d_i : distance between coordinates of the i^{th} atom in the two structures.¹⁵⁹

RMSD analysis was employed to compare the initial protein structure (docked conformation) and the final equilibrated structure. The RMSD between the structures is calculated at every frame of the 2500 MD simulation frames.

The frequency of the contact between the ligand and surrounding protein residues was calculated for the resulting MD trajectory to see how the protein-ligand interaction scheme changes from the predicted docked pose (initial conformation of MD simulation). In the present study, the contact between the ligand and the residue is defined when the distance between any of the ligand atoms and any of the residue atoms is closer than 3.0 Å.

2.6 In-vitro screening of predicted anti-androgen responses

The response of AR mutants to increasing concentrations of various anti-androgens has been measured using a luciferase-reporter based transcriptional assay. PC3 cells lacking the AR were transiently co-transfected with 25 ng of either the wild-type or a mutated form of the AR and 25 ng of the reporter plasmid pARR3-tk-luciferase.⁹ 48 hours after transfection, cells were stimulated with 0.1 nM of the synthetic androgen R1881 and treated with 0.1 % DMSO (for the control) or increasing concentrations of anti-androgens (Bicalutamide, Enzalutamide, and Hydroxyflutamide). Cells were lysed after 24 hours and the luciferase activity was quantified. Each concentration was assayed in quadruplicate $n = 4$, with 2 to 3 biological replicates. Results were normalized to the wild-type AR activity.

Chapter 3: Cheminformatics modeling of AR mutant— drug responses

The Cheminformatics modeling pipeline as described in Figure 2.2 was adopted for generating and *in silico* evaluating AR mutants and predicting their drug responses.

3.1 In-silico AR mutant analyses

The *in-silico* AR mutant analyses provided valuable insights in understanding and differentiating between wild-type and mutant-AR receptors. These modeling experiments revealed several striking findings that may correspond to structural changes in the receptor, potentially describing the protein-ligand interactions that dictate the biological response and viability.

3.1.1 Generating AR mutant structures

A total of 25 training set instances were created, combining 24 reported AR mutants along with the wild-type AR as described in Section 2.1.1. The Residue Scan module of MOE was used to generate AR mutant structures. This module provides the functionality of site-directed mutagenesis for a maximum of 6 mutation sites or substitution by 20 amino acids at 6 sites enumerating to about 64 million sequences. To construct the test set 28 residues in the ABS were substituted by 19 other residues, producing a total of 532 mutant structure models. The mutated residues have been listed in Table 3.1. 22 out of the generated test-set structure models were filtered out in the protein preparation step. Receptor grids were then generated for the remaining 25 training and 510 test set structures (535 in total). The ligand molecules were docked into the ABS, using Glide XP mode. Compared to Glide SP, higher amount of computational resources were consumed. For instance a SP docking job to dock 7 ligand molecules to 1 protein structure consume 3 minutes, over the localhost whereas, the same job done through XP mode would take up to 18 minutes to incorporate the docking results.

S. No.	Residue #	Mutated Residue
1	685	V
2	702	L
3	705	L
4	706	N
5	708	L
6	709	G
7	712	Q
8	716	V
9	739	Q
10	742	W
11	743	M
12	746	M
13	747	V
14	750	M
15	765	F
16	766	A
17	781	M
18	784	Q
19	788	M
20	873	E
21	874	L
22	877	L
23	878	T
24	879	F
25	881	L
26	892	F
27	896	M
28	900	I

Table 3.1: Residues of the wildtype AR (WT-AR) mutated to engineer in-silico mutants

3.1.2 Structure-based analyses

The molecular docking revealed several distinct binding poses that exhibits a clear contrast between the wild type and mutant AR receptor structures. The ABS is a well buried pocket, hydrophobic in nature that has been linked to the numerous hydrophobic side chains that build up the ABS.⁴¹ The side chains make several contacts with the ABS binders through Van der Waals, and electrostatic interactions. Interestingly, the ABS also contains two polar regions

located at the opposite ends: with R753 on one, and N706 on the other. This region can be thought of as a ‘hook’ region, that helps in anchoring the steroidal ligands.²³

The crystal structures of the AR LBD reported all represent the protein in an agonist conformation. The AR-LBD antagonist crystal structure is unavailable till date. Investigations into the receptor-ligand interactions of the wild type versus mutants possible through computational modeling provide substantial clues that could plausible explanations for the antagonist to agonist response switch.

In 2013, *Gao et al.* reported a new mutant F877L that agonizes the then recently approved anti-androgen Enzalutamide. Retrospective analysis to study the interaction of Enzalutamide within the wild-type ABS, suggests that the binding conformation can be correlated with an agonist (AR growth activating)/antagonist (AR growth inhibiting) response. The expected response of an anti-androgen is to inhibit the proliferation of the cells of the prostate (antagonist) however, in certain cases this anticipated behavior is not observed given promiscuity in the mutated AR yielding agonist response, promoting cell growth. In the docking experiment, Enzalutamide was found to bind in a distinct conformational orientation as compared to WT— Enzalutamide complex, as shown in Figure 3.1. This was accompanied by elevated levels of protein-ligand interactions in the F877L—Enzalutamide complex. (See Figure 1.3) The substitution of a larger phenylalanine side chain by a smaller leucine, results in an increased ligand accessible surface area. This permits Enzalutamide to traverse deep into the ABS, making some key interactions such as π - π stacking interactions between the ring moiety with the benzene ring of the F765 residue in addition to other Hydrogen bond interactions with residue backbones, especially with the L877 residue.

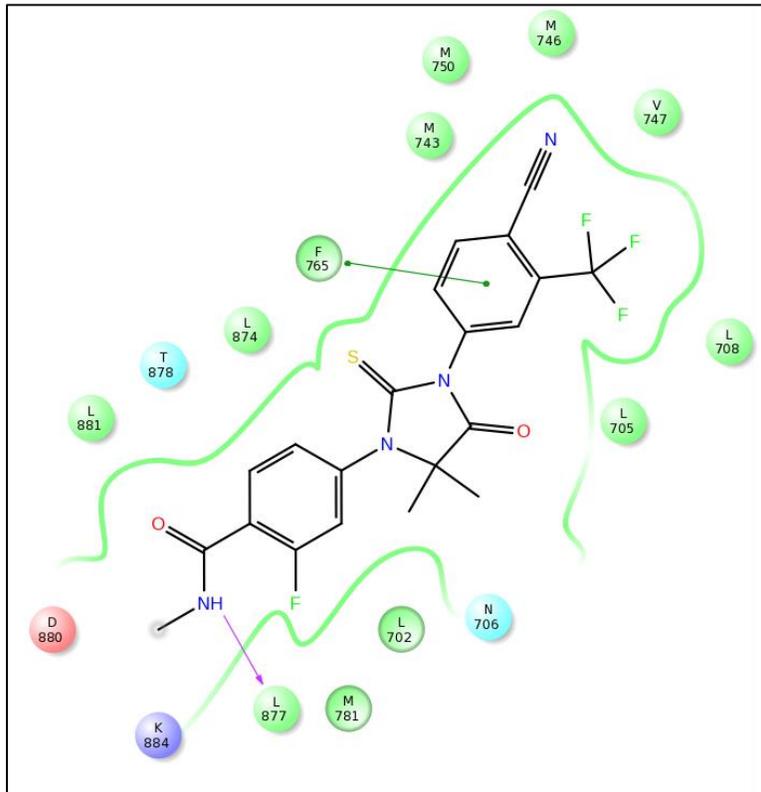


Figure 3.1: Ligand interaction diagram-Enzalutamide bound to the androgen binding site (ABS) of F877L mutant, L877 residue interacts with the ligand molecule through H-bonding in addition to π - π stacking between benzene ring of F765 and ring moiety of Enzalutamide

Overall, contrasting with the interactions of the wild-type receptor with the native as well as other ligands opened up a new avenue for further analysis. When these aforementioned receptor-ligand interactions were quantified, a correlation between the computed values and functional characterization could also be established.

3.2 Molecular descriptors and statistical trends

A total of 225 4D-Inductive descriptors along with Glide Per-Residue Interaction scores were calculated for the protein-ligand complexes. Glide per-residue interaction scores for residues lying within a 10 Å distance of the grid center were calculated. The Coulomb, van der Waals,

and H-bonding scores were computed and written as structure-level properties for each ligand to the Maestro file generated.

3.2.1 List of 4D-inductive descriptors computed

The following 4D-Inductive descriptors were developed in SVL for this study, where R is the receptor and L is the ligand. The list contains 84 descriptors, and the rest generated were residue-specific iterations of the same Rs, Abs_Rs and Sigma_L_R descriptors, totaling up to 225 along with Glide per-residue interaction scores.

<i>S. No.</i>	<i>Name</i>	<i>Description</i>
1.	Rs_L_R	Steric effect between the ligand and 10 Å region of the receptor
2.	Abs_Rs_L_R	Absolute steric effect between the ligand and the receptor (10 Å)
3.	Sigma_L_R	Cumulative sum of interactions between the ligand and receptor
4.	Q_L_R	Overall charge interaction between ligand and receptor
5.	R_Csp3-L_Csp3	Steric interaction between Carbon sp ³ of L and R
6.	R_Csp3-L_Csp2	Steric interaction between Carbon sp ³ of R and sp ² carbon of L
7.	R_Csp3-L_Csp	Steric interaction between carbon sp ³ of R and sp carbon of L
8.	R_Csp3-L_Osp3	Steric interaction between carbon sp ³ of R and oxygen sp ³ of L
9.	R_Csp3-L_Osp2	Steric interaction between carbon sp ³ of R and oxygen sp ² of L
10.	R_Csp3-L_Nsp3	Steric interaction between carbon sp ³ of R and nitrogen sp ³ of L
11.	R_Csp3-L_Nsp2	Steric interaction between carbon sp ³ of R and nitrogen sp ² of L
12.	R_Csp3-L_Nsp	Steric interaction between carbon sp ³ of R and nitrogen sp of L
13.	R_Csp3-L_Ssp3	Steric interaction between carbon sp ³ of R and sulphur sp ³ of L
14.	R_Csp3-L_Ssp2	Steric interaction between carbon sp ³ of R and sulphur sp ² of L

<i>S. No.</i>	<i>Name</i>	<i>Description</i>
15.	R_Csp3-L_Fsp3	Steric interaction between carbon sp ³ of R and fluorine sp ³ of L
16.	R_Csp2-L_Csp3	Steric interaction of carbon sp ² of R and carbon sp ³ of L
17.	R_Csp2-L_Csp2	Steric interaction of carbon sp ² of R and carbon sp ² of L
18.	R_Csp2-L_Csp	Steric interaction of carbon sp ² of R and carbon sp of L
19.	R_Csp2-L_Osp3	Steric interaction between carbon sp ² of R and oxygen sp ³ of L
20.	R_Csp2-L_Osp2	Steric interaction between carbon sp ² of R and oxygen sp ² of L
21.	R_Csp2-L_Nsp3	Steric interaction between carbon sp ² of R and nitrogen sp ³ of L
22.	R_Csp2-L_Nsp2	Steric interaction between carbon sp ² of R and nitrogen sp ² of L
23.	R_Csp2-L_Nsp	Steric interaction between carbon sp ² of R and nitrogen sp of L
24.	R_Csp2-L_Ssp3	Steric interaction between carbon sp ² of R and sulphur sp ³ of L
25.	R_Csp2-L_Ssp2	Steric interaction between carbon sp ² of R and sulphur sp ² of L
26.	R_Csp2-L_Fsp3	Steric interaction between carbon sp ² of R and fluorine sp ³ of L
27.	R_Osp3-L_Csp3	Steric interaction between oxygen sp ³ of R and carbon sp ³ of L
28.	R_Osp3-L_Csp2	Steric interaction between oxygen sp ³ of R and carbon sp ² of L
29.	R_Osp3-L_Csp	Steric interaction between oxygen sp ³ of R and carbon sp of L
30.	R_Osp3-L_Osp3	Steric interaction between oxygen sp ³ of R and oxygen sp ³ of L
31.	R_Osp3-L_Osp2	Steric interaction between oxygen sp ³ of R and oxygen sp ² of L
32.	R_Osp3-L_Nsp3	Steric interaction between oxygen sp ³ of R and nitrogen sp ³ of L
33.	R_Osp3-L_Nsp2	Steric interaction between oxygen sp ³ of R and nitrogen sp ² of L
34.	R_Osp3-L_Nsp	Steric interaction between oxygen sp ³ of R and nitrogen sp of L
35.	R_Osp3-L_Ssp3	Steric interaction between oxygen sp ³ of R and sulphur sp ³ of L
36.	R_Osp3-L_Ssp2	Steric interaction between oxygen sp ³ of R and sulphur sp ² of L

<i>S. No.</i>	<i>Name</i>	<i>Description</i>
37.	R_Osp3-L_Fsp3	Steric interaction between oxygen sp ³ of R and fluorine sp ³ of L
38.	R_Osp2-L_Csp3	Steric interaction between oxygen sp ² of R and carbon sp ³ of L
39.	R_Osp2-L_Csp2	Steric interaction between oxygen sp ² of R and carbon sp ² of L
40.	R_Osp2-L_Csp	Steric interaction between oxygen sp ² of R and carbon sp of L
41.	R_Osp2-L_Osp3	Steric interaction between oxygen sp ² of R and oxygen sp ³ of L
42.	R_Osp2-L_Osp2	Steric interaction between oxygen sp ² of R and oxygen sp ² of L
43.	R_Osp2-L_Nsp3	Steric interaction between oxygen sp ² of R and nitrogen sp ³ of L
44.	R_Osp2-L_Nsp2	Steric interaction between oxygen sp ² of R and nitrogen sp ² of L
45.	R_Osp2-L_Nsp	Steric interaction between oxygen sp ² of R and nitrogen sp of L
46.	R_Osp2-L_Ssp3	Steric interaction between oxygen sp ² of R and sulphur sp ³ of L
47.	R_Osp2-L_Ssp2	Steric interaction between oxygen sp ² of R and sulphur sp ² of L
48.	R_Osp2-L_Fsp3	Steric interaction between oxygen sp ² of R and fluorine sp ³ of L
49.	R_Nsp3-L_Csp3	Steric interaction between nitrogen sp ³ of R and carbon sp ³ of L
50.	R_Nsp3-L_Csp2	Steric interaction between nitrogen sp ³ of R and carbon sp ² of L
51.	R_Nsp3-L_Csp	Steric interaction between nitrogen sp ³ of R and carbon sp of L
52.	R_Nsp3-L_Osp3	Steric interaction between nitrogen sp ³ of R and oxygen sp ³ of L
53.	R_Nsp3-L_Osp2	Steric interaction between nitrogen sp ³ of R and oxygen sp ² of L
54.	R_Nsp3-L_Nsp3	Steric interaction between nitrogen sp ³ of R and nitrogen sp ³ of L
55.	R_Nsp3-L_Nsp2	Steric interaction between nitrogen sp ³ of R and nitrogen sp ² of L
56.	R_Nsp3-L_Nsp	Steric interaction between nitrogen sp ³ of R and nitrogen sp of L
57.	R_Nsp3-L_Ssp3	Steric interaction between nitrogen sp ³ of R and sulphur sp ³ of L

<i>S. No.</i>	<i>Name</i>	<i>Description</i>
		L
58.	R_Nsp3-L_Ssp2	Steric interaction between nitrogen sp3 of R and sulphur sp2 of L
59.	R_Nsp3-L_Fsp3	Steric interaction between nitrogen sp3 of R and fluorine sp3 of L
60.	R_Nsp2-L_Csp3	Steric interaction between nitrogen sp2 of R and carbon sp3 of L
61.	R_Nsp2-L_Csp2	Steric interaction between nitrogen sp2 of R and carbon sp2 of L
62.	R_Nsp2-L_Csp	Steric interaction between nitrogen sp2 of R and carbon sp of L
63.	R_Nsp2-L_Osp3	Steric interaction between nitrogen sp2 of R and oxygen sp3 of L
64.	R_Nsp2-L_Osp2	Steric interaction between nitrogen sp2 of R and oxygen sp2 of L
65.	R_Nsp2-L_Nsp3	Steric interaction between nitrogen sp2 of R and nitrogen sp3 of L
66.	R_Nsp2-L_Nsp2	Steric interaction between nitrogen sp2 of R and nitrogen sp2 of L
67.	R_Nsp2-L_Nsp	Steric interaction between nitrogen sp2 of R and nitrogen sp of L
68.	R_Nsp2-L_Ssp3	Steric interaction between nitrogen sp2 of R and sulphur sp3 of L
69.	R_Nsp2-L_Ssp2	Steric interaction between nitrogen sp2 of R and sulphur sp2 of L
70.	R_Nsp2-L_Fsp3	Steric interaction between nitrogen sp2 of R and fluorine sp3 of L
71.	R_Ssp3-L_Csp3	Steric interaction between sulphur sp3 of R and carbon sp3 of L
72.	R_Ssp3-L_Csp2	Steric interaction between sulphur sp3 of R and carbon sp2 of L
73.	R_Ssp3-L_Csp	Steric interaction between sulphur sp3 of R and carbon sp of L
74.	R_Ssp3-L_Osp3	Steric interaction between sulphur sp3 of R and oxygen sp3 of L

<i>S. No.</i>	<i>Name</i>	<i>Description</i>
75.	R_Ssp3-L_Osp2	Steric interaction between sulphur sp3 of R and oxygen sp2of L
76.	R_Ssp3-L_Nsp3	Steric interaction between sulphur sp3 of R and nitrogen sp3of L
77.	R_Ssp3-L_Nsp2	Steric interaction between sulphur sp3 of R and nitrogen sp2of L
78.	R_Ssp3-L_Nsp	Steric interaction between sulphur sp3 of R and nitrogen sp of L
79.	R_Ssp3-L_Ssp3	Steric interaction between sulphur sp3 of R and sulphur sp3of L
80.	R_Ssp3-L_Ssp2	Steric interaction between sulphur sp3 of R and sulphur sp2of L
81.	R_Ssp3-L_Fsp3	Steric interaction between sulphur sp3 of R and fluorine sp3of L
82.	AA#_Rs_AA_L	Steric interaction between amino acid residue AA# with the ligand
83.	AA#_Abs_Rs_AA_L	Absolute steric interaction between amino acid residue AA# with the ligand
84.	AA#_Sigma_AA_L	Overall charge interaction between amino acid residue AA# with the ligand

Table 3.2: List of 4D-inductive descriptors computed, AA# descriptors were calculated for the mutated residues

3.2.2 Drug response correlation with molecular descriptor values

Lallous et al. reported that the AR mutant H875Y has one of the most aggressive agonist responses to all currently available anti-androgens.⁸ Partial agonist response represents the biphasic behavior of the anti-androgen that would initially activate the receptor followed by a decline in activation and re-activation at higher concentrations. Whereas, when the receptor is activated by lower concentrations of the anti-androgens and the trend is consistent with rising concentrations can be inferred to as complete agonist response. Hydroxyflutamide acts as a complete agonist against this mutant. This behavior can be traced to the descriptor values

computed for H875Y mutant complex with Hydroxyflutamide. Attributed to its smaller surface area, Hydroxyflutamide does not encounter as much steric hindrance as compared to Bicalutamide and Enzalutamide, permitting more chances to make more contacts with different residues within the mutant pocket. Furthermore, it has been reported that this mutant yields weaker agonist responses to Bicalutamide and Enzalutamide and can be observed in the descriptor values that signify the speculated correlation of the drug responses and protein-ligand interactions.

<i>H875Y</i>	<i>Rs_L_R</i>	<i>Abs_Rs_L_R</i>	<i>875 Rs_AA_L</i>	<i>R_Csp3-L_Fsp3</i>	<i>R_Nsp2-L-Nsp</i>	<i>Response</i>
<i>Bicalutamide</i>	239.39	70.57	0.37	13.37	1.17	Partial Agonist
<i>Enzalutamide</i>	268.38	78.94	0.48	13.12	1.16	Partial Agonist
<i>Hydroxyflutamide</i>	157.73	46.65	0.19	9.84	0.00	Complete Agonist

Table 3.3: Descriptor value correlation with biological response obtained for mutant H875Y

3.2.3 Descriptor pruning and ranking

To analyze the spread and distribution of descriptor values, density measure was used through ggplot2 package in R.¹⁶⁰ The attributes were pruned and prioritized through Boruta and caret packages of R programming language and the top ranking attributes were then chosen for QSAR modeling. After removal of any correlation with nominal attribute (activity, in this case), 3 shadow attributes (shown in Blue in Figure 3.2) were added to the dataset, and Z-scores were accumulated through random forest algorithm implementation. The list of all the 225 attributes (in order of their importance as shown in figures 3.2 and 3.3) can be found in Appendix B.

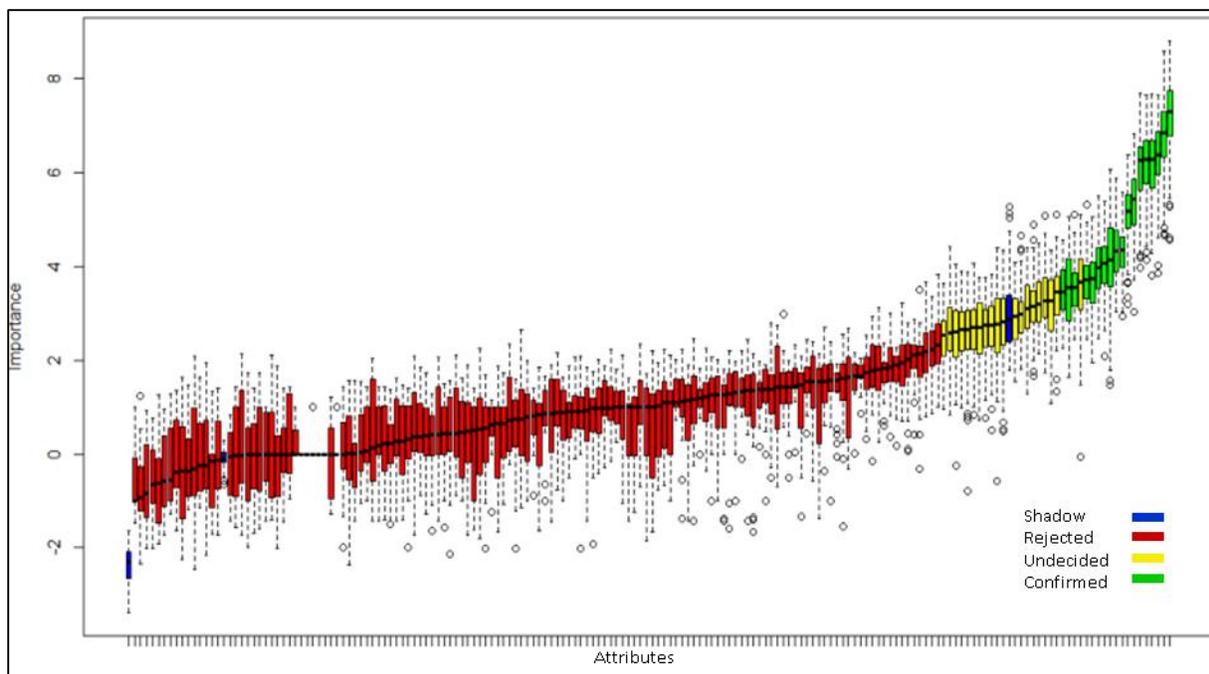


Figure 3.2: Prioritization of attributes, initial run of Boruta package. Shadow attributes: blue, Rejected attributes: red, Undecided importance of attributes: yellow, Confirmed importance: green

Based upon the MZSA score (Maximum Z-Score of Shadow Attribute), attributes scoring were sorted. In the very first run, through 99 iterations performed over 19.66 seconds, Boruta segregated the attributes into 3 categories of importance (confirmed, undecided, rejected).

135 attributes were rejected after this run, whereas 18 were confirmed to be important for model building. Within the undecided category, 20 attributes were present, which had a tentative ‘undecided importance’. In the second Boruta run, the undecided class was re-evaluated, and 28 attributes were confirmed to be important. (See Figure 3.3)

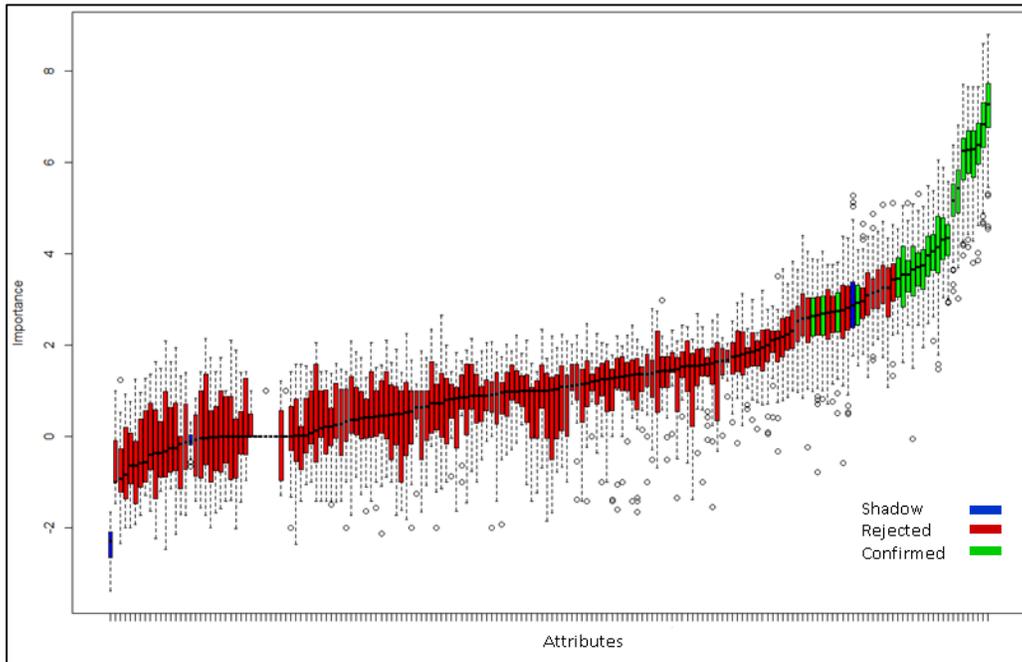


Figure 3.3: Attribute prioritization, final run- segregates into confirmed (green) and rejected (red) classes based on importance

The attribute prioritization output enlisting the attributes and the importance parameters have been listed in Table 3.4.

<i>S.No.</i>	<i>Attribute</i>	<i>meanImp</i>	<i>medianImp</i>	<i>minImp</i>	<i>maxImp</i>	<i>normHits</i>	<i>Decision</i>
1	X896.Abs_Rs_AA_L	7.16	7.39	4.40	8.89	1.00	Confirmed
2	X716.Rs_AA_L	5.88	5.92	3.43	7.56	1.00	Confirmed
3	X896.Rs_AA_L	5.35	5.56	3.09	6.98	0.99	Confirmed
4	R_Osp3.L_Nsp3	5.24	5.30	3.29	6.89	0.98	Confirmed
5	R_Osp3.L_Csp2	5.16	5.26	3.59	6.49	0.98	Confirmed
6	X731.Rs_AA_L	4.87	4.96	3.24	6.35	0.97	Confirmed
7	X731.Abs_Rs_AA_L	4.58	4.63	2.38	6.24	0.93	Confirmed
8	X742.Rs_AA_L	4.74	4.73	2.84	6.19	0.96	Confirmed

<i>S.No.</i>	<i>Attribute</i>	<i>meanImp</i>	<i>medianImp</i>	<i>minImp</i>	<i>maxImp</i>	<i>normHits</i>	<i>Decision</i>
9	r_i_glide_emodel	4.34	4.50	1.69	6.18	0.87	Confirmed
10	R_Osp2.L_Nsp3	4.83	4.81	3.12	6.13	0.96	Confirmed
11	R_Osp3.L_Fsp3	4.56	4.57	2.71	5.88	0.93	Confirmed
12	R_Nsp3.L_Ssp2	4.11	4.31	1.73	5.83	0.80	Confirmed
13	r_glide_res.871_Eint	4.34	4.54	1.48	5.81	0.88	Confirmed
14	X742.Abs_Rs_AA_L	4.21	4.20	2.51	5.72	0.90	Confirmed
15	R_Csp2.L_Nsp3	4.38	4.38	2.87	5.68	0.92	Confirmed
16	X716.Abs_Rs_AA_L	4.20	4.23	2.89	5.51	0.92	Confirmed
17	R_Csp3.L_Nsp3	3.93	3.96	2.23	5.47	0.87	Confirmed
18	r_glide_res.870_Eint	3.95	3.95	2.46	5.40	0.88	Confirmed
19	r_glide_res.878_dist	3.69	3.80	0.89	5.35	0.76	Confirmed
20	r_glide_res.779_Eint	3.71	3.79	1.71	5.30	0.74	Confirmed
21	R_Osp3.L_Csp3	3.65	3.63	2.43	5.25	0.78	Confirmed
22	X898.Abs_Rs_AA_L	3.64	3.64	2.03	5.10	0.77	Confirmed
23	R_Csp3.L_Nsp	3.11	3.30	0.38	5.03	0.57	Confirmed
24	R_Osp2.L_Csp2	3.75	3.76	2.21	5.03	0.78	Confirmed
25	X894.Abs_Rs_AA_L	3.81	3.92	1.92	5.01	0.81	Confirmed
26	X882.Sigma_AA_L	3.53	3.57	1.66	4.91	0.71	Confirmed
27	R_Csp2.L_Csp2	3.63	3.70	2.07	4.86	0.74	Confirmed
28	R_Ssp3.L_Nsp3	3.65	3.65	2.20	4.75	0.74	Confirmed

Table 3.4: List of molecular descriptors with confirmed importance as per Boruta implementation for attribute prioritization

3.3 QSAR model development and validation

Prior to QSAR model development, lowest number of attributes are required to be selected that would yield maximum performance as well as lowest error rate. This step controls the possible redundancy scenario. RFE (recursive feature elimination) was employed via the caret package to

determine the number of attributes required to build up QSAR models with minimum RMSE (root mean squared error). See Figure 3.4

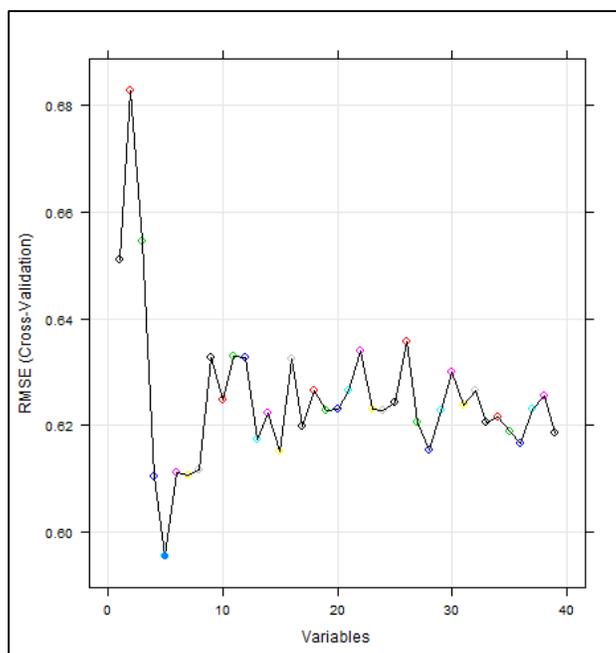


Figure 3.4: Determining the number of attributes to be used for QSAR modeling

A total of 5 attributes were used for constructing the QSAR models, as shown by the caret implementation of RFE, to yield lowest RMSE upon cross-validation shown in Table 3.5.

<i>No.</i>	<i>Name</i>	<i>Definition</i>	<i>Class</i>
1.	R_Csp3-L_Nsp3	Steric hindrance on sp ³ hybridized Carbons of receptor (within 10 Å of the ligand) by sp ³ hybridized Nitrogen of ligand	4D-Inductive
2.	R_Csp2-L_NSp3	Steric hindrance on sp ² hybridized Carbons of receptor (within 10 Å of the ligand) by sp ³ hybridized Nitrogen of ligand	4D-Inductive
3.	R_Osp3-L_Nsp3	Steric hindrance on sp ³ hybridized Oxygens of receptor	4D-

<i>No.</i>	<i>Name</i>	<i>Definition</i>	<i>Class</i>
		(within 10 Å of the ligand) by sp ³ hybridized Nitrogen of ligand	Inductive
4.	R_Osp2-L_Nsp3	Steric hindrance on sp ² hybridized Oxygens of receptor (within 10 Å of the ligand) by sp ³ hybridized Nitrogen of ligand	4D-Inductive
5.	r_i_glide_emodel	Weighting of force field components (electrostatic and van der Waals energies) for picking best ligand binding pose	Glide

Table 3.5: Attributes selected for QSAR model construction

3.3.1 Categorical QSAR models predict biological responses

QSAR models were applied for screening the therapeutic biological responses of anti-androgens with respect to various mutants. The nominal attribute ‘activity’ was amalgamated with the computed molecular descriptors. This attribute was created based on the observations of the functional characterization experiment performed at the Vancouver Prostate Centre, populating instances with either +1 (agonist) or -1 (antagonist) classes.⁸

Similarly, another dichotomous attribute, ‘active’ was added to the training set, that denotes the activation or inactivation of the receptor by native ligand, DHT. This attribute was populated with binary values 1 (active) or 0 (inactive). Briefly, categorical QSAR models were then generated for predicting values of the aforementioned nominal attributes, by using a collection of 7 machine learning classification algorithms implemented in WEKA, described in Section 1.3.2. A 10-fold cross validation initially validated the train-model, followed by external test

validation. The predictions were then screened by a consensus voting protocol to determine the final predicted outcome for each of the instances.

3.3.2 Significance of consensus voting for classification algorithms

The classifier algorithms incorporated into the QSAR modeling pipeline contain a variety of weak and strong learners. Boosting algorithms, i.e. combining weak learners to produce a strong learner, may be highly effective in solving the problem. Although, the downside of boosting is that, it may lead to overfitting. The performance may seem to be enhanced however this could be rendered to the over-generalization based on the training dataset.¹⁶¹ Therefore, by a consensus voting approach, we were able to eliminate most of the probable causes affecting model performance, eliminating any algorithmic biasness. Each instance in the dataset was a protein-ligand complex such as F877L-Enzalutamide, F877L-Bicalutamide etc. The responses predicted by individual algorithms for all instances were populated and the response, inferred by the majority of algorithms was considered as the final response value.

3.3.3 Statistical accuracy of QSAR predictions

The statistical parameters of sensitivity, specificity, PPV etc. indicate the performance of a QSAR model. The relevance and significance of the predictions generated by QSAR models largely depend on how accurately they were identified and validated.

In order to validate the agonist-antagonist QSAR model, the 10-fold cross-validation approach employed, provide significantly high measures of sensitivity, specificity, PPV and overall accuracy (as shown in Table 3.6). High sensitivity and specificity measures indicate accurate assessment of both agonist and antagonist responses. The overall accuracy was over 82% for all different QSAR models operated upon for model development.

<i>S. No.</i>	<i>Algorithms</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>PPV</i>	<i>Accuracy</i>	<i>AUC of ROC</i>
1	<i>Bagging</i>	0.86	0.87	0.90	86.90%	0.89
2	<i>Dagging</i>	0.75	0.93	0.95	82.14%	0.89
3	<i>DecisionStump</i>	0.90	0.90	0.93	86.90%	0.82
4	<i>IBk</i>	0.79	0.87	0.91	82.14%	0.85
5	<i>LibSVM</i>	0.86	0.90	0.93	88.09%	0.88
6	<i>OneR</i>	0.86	0.83	0.90	85.71%	0.85
7	<i>RandomForest</i>	0.86	0.87	0.92	86.90%	0.87

Table 3.6: Performance statistics of QSAR classification models

Additionally, the receiver operating characteristic (ROC) curves compared sensitivity (TPR) vs 1-specificity (FPR) across the range of predicted outcomes (See Figure 3.5) across a range of threshold values ranging between 0 and 1. Initially, the outcomes were analyzed when the threshold for finding true positives was set as 0 in this case all the instances were predicted as positives until gradually increasing the value till 1, to rigorously evaluate each of the different classifiers performance at varying threshold values. The ClassifierPerformanceEvaluator module of WEKA's KnowledgeFlow suite was used to evaluate the performance of the training dataset.¹⁰⁸ The area under the ROC curve, gives a measure of the model's performance. Approximately 85% of the area was covered under the ROC curve, which exemplifies an effective measure of accuracy that can be considered to obtain meaningful interpretations of the predicted responses.⁹ A consensus voting approach was applied for the predictions, to enhance the accuracy.

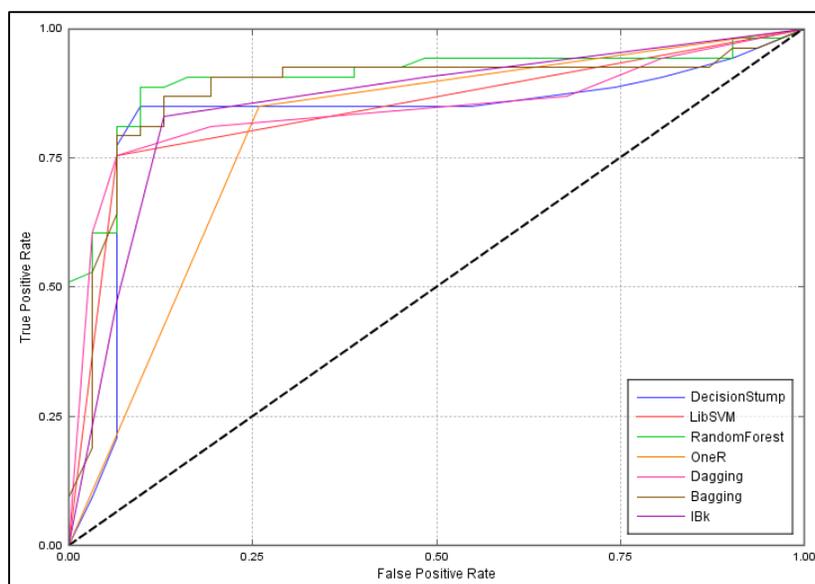


Figure 3.5: Over 85% of area under the ROC curve, demonstrates the high diagnostic ability, effectively discriminating true positives from the false positives

3.3.4 Applicability domain assessment

The descriptor space for both train and test sets was evaluated between the minimum and maximum descriptor values ($\pm 15\%$). The selection of the $\pm 15\%$ threshold reflects the standard error value that is used for leverage based applicability domain assessment.¹⁰⁰ The presumption of this procedure is that predictive accuracy for those instances that lie within the range of descriptor values ($\pm 15\%$) tend to be higher compared to those that are beyond the permissible threshold. The values for the 5 molecular descriptors that were used for QSAR model building were well within the applicability domain. In Figure 3.6, red arcs represent the spread of the test set values, which are within the range of the training set values (green arcs) $\pm 15\%$. The minimum values have been shown in boxes on the left and maximum values in the boxes on the right.

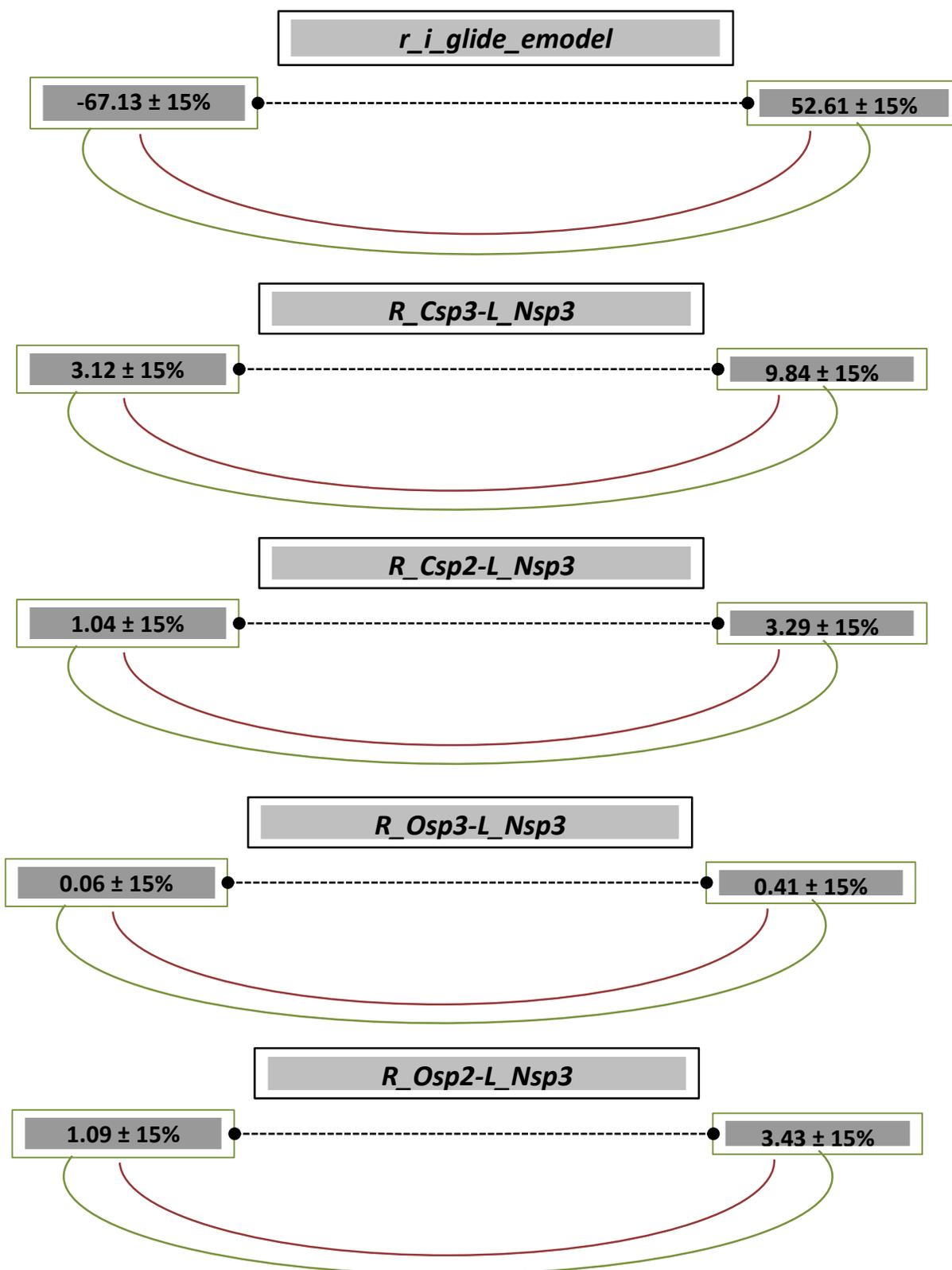


Figure 3.6: Applicability domain assessment of molecular descriptor values, utilized in QSAR modeling.

Red arcs represent the spread of the Test set values, which are within the range of the Training set values

(Green arcs) ± 15%.

3.3.5 Predicted mutants

Upon consensus voting of the QSAR model predictions for the test-set, a total of 12 mutants were predicted to produce an agonist effect against the currently used anti-androgens.

<i>S. No.</i>	<i>Mutant</i>
1	R753H
2	L874N
3	F877A
4	F877G
5	R753P
6	R753S
7	F765G
8	T878G
9	F765Q
10	L705W
11	T878V
12	F877T

Table 3.7: AR-LBD mutants predicted to yield an agonist response towards anti-androgens

The 12 mutants shown in Table 3.7 were experimentally created and tested (by the procedure described in Section 2.6) for their biological activation using a R1881 stimulus followed by evaluation of their response to the anti-androgens. Unfortunately, 9 of these mutants were transcriptionally inactive and no biological activity whatsoever. This was an interesting finding, since modification in those residues that help in the binding of anti-androgens results in structural disruption of AR making it non-viable for survival.

For instance, molecular docking shows that the benzene ring of the residue F765 plays an instrumental role in harboring the anti-androgens within the ABS pocket. When this residue was substituted by a much smaller residue such as glycine, it resulted in complete structural disruption.

This information could not be captured even through MD simulation production runs which initially were of 10 ns. However computationally expensive, longer MD production runs could provide better estimation of the receptor's stability in its native environment and overall system equilibrium. The production run duration was increased up to 25 ns. Two mutants F877G and F877T exhibited biological activity but with a very low signal intensity and hence, could not be further investigated. The mutant T878G was biologically viable and initial experiments showed the activation by DHT.

3.4 Molecular dynamics simulations analysis

The major objective of performing MD simulations was to ascertain the binding fitness of the ligand and more importantly, the stability of the receptor structure. Additionally, contact frequency analysis provided a retrospective outlook of interaction modifications that could potentially explain observed biological activity. The MD simulations were carried out on Westgrid's high performance computing cluster (HPCC). This involved several steps as described in Section 2.5.

The energy equilibrated structures offer comprehension of the modifications the receptor-ligand complex undergoes when subjected to physical conditions of changed pressure, heat and solvation. The digression from the rigid structure obtained from molecular docking denotes the flexibility of the receptor structure, as well as the altered binding conformations of the ligand. The modified thermodynamic properties calculated during the simulations can be correlated to the biological activity such as agonist.

3.4.1 Protein structural stability evaluation

The equilibrium of the receptor – ligand system was analyzed through RMSD calculations. The RMSD between the initial and equilibrated structures was computed. The time duration of the MD simulation i.e. 25 ns has been shown on x-axis of the RMSD graph in Figure 3.7 and RMSD in Å has been shown on y-axis. The equilibrium fluctuations can be seen as different on modes on the graph, these fluctuations reach a constant value when the system has been equilibrated, denoted by minor RMSD differences. The steady state values attained during the simulation indicate the stability.¹⁶² The fluctuations correspond to the flexibility and induced fit changes of ligand binding within the protein pocket, when the system is exposed to ambient conditions.⁸³ For the wild-type (WT) AR- Enzalutamide complex, the average RMSD for the backbone was calculated as 1.16 Å in reference to initial docked structure, whereas for the Enzalutamide molecule the RMSD was 1.87 Å. The equilibrated structure validates the obtained docking pose, through its consistency throughout the simulation time. Also, the ligand molecule did not escape the binding pocket upon solvent exposure. The different WT complexes as with DHT, Enzalutamide etc. followed a very stable trajectory with no large deviations. The binding poses of the ligands remained unchanged as well as their alignment within the pocket was consistent, that validates the docked poses obtained. (See Figure 3.7)

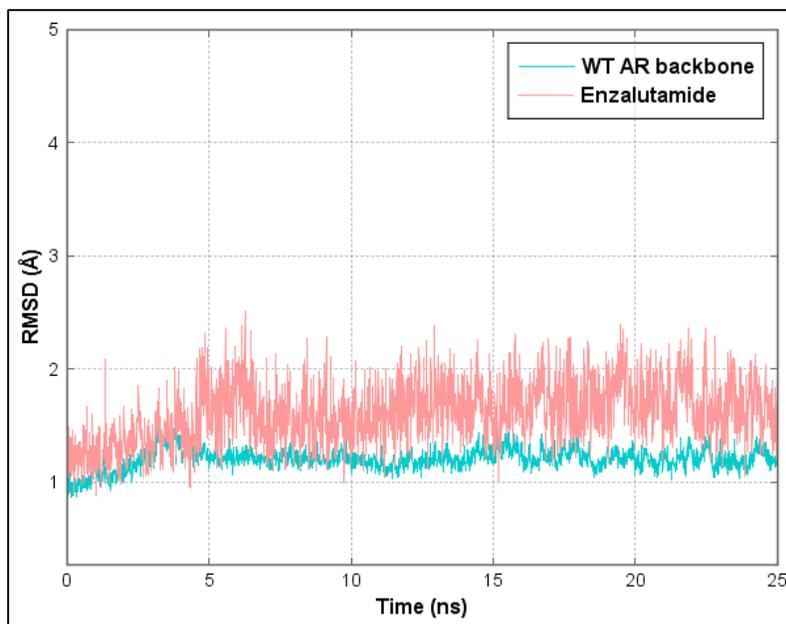


Figure 3.7: Structural stability of the WT AR - Enzalutamide complex illustrated through a stable trajectory obtained for 25ns MD simulation

On the other hand, for the double mutant F877L/T878A-Enzalutamide complex exhibited a dissimilar case as compared to the WT- Enzalutamide system. A large deviation from the initial docked pose could be observed in the obtained MD trajectory. This modification corresponds to the altered conformation of Enzalutamide binding (RMSD 2.45 Å) within the ABS pocket of the mutant, during the course of the production run. Enzalutamide undergoes a conformational change at 17.4 ns during the equilibration run. The conformation change of Enzalutamide can also be a plausible explanation of the biological response observed in case of this double mutant which is agonized by Enzalutamide opposite to the anticipated outcome. (See Figure 3.8)

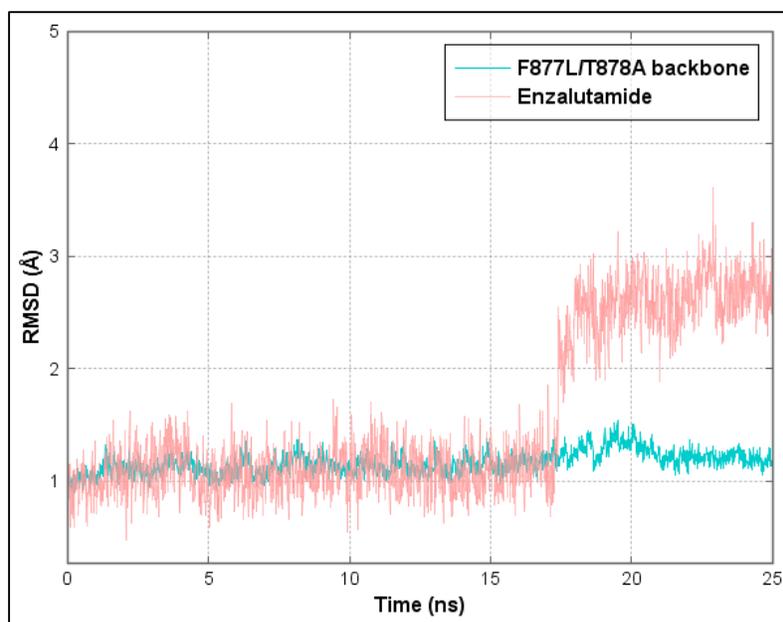


Figure 3.8: F877L/T878A - Enzalutamide complex trajectory obtained upon 25ns MD simulation

Table 3.8 enlists some of the RMSD calculated for AR- ligand complexes.

<i>S. No.</i>	<i>Complex</i>	<i>Backbone (Å)</i>	<i>Ligand (Å)</i>
1.	WT-DHT	1.35	1.42
2.	WT-Enzalutamide	1.16	1.87
3.	F877L-Enzalutamide	1.24	1.42
4.	T878A-Enzalutamide	1.11	1.13
5.	F877L/T878A-Enzalutamide	1.15	2.45

Table 3.8: RMSD comparison between initial docked structures and energy equilibrated MD structures

3.4.2 Contact frequency analysis

The contact frequency analysis, for the F877L/T878A Enzalutamide complex evaluates the frequency statistics for the contacts maintained between the receptor and the ligand over the 25 ns simulation time. The consistency of the initial contacts, generated through molecular docking and their changes over the course of MD run, can provide insights into how the ligand-receptor

interaction can vary. One of the most critical contacts was the backbone interaction of L705 with the ligand molecule that was consistent through 2498 of the 2500 simulation steps (99.92%). This interaction along with the A878 (99.6%) and F892 (94.16%) indicate Enzalutamide's binding in an agonist conformation yielding the corresponding response. (See Figure 3.9)

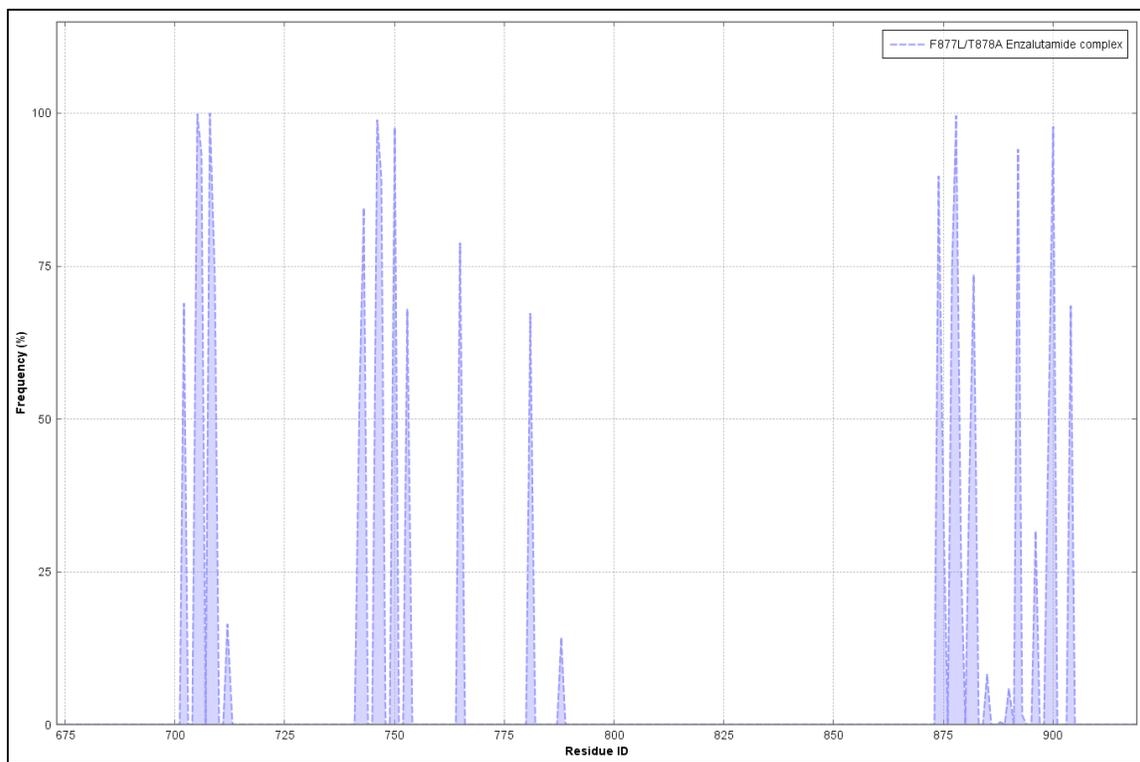


Figure 3.9: Contact frequency analysis of F877L/T878A - Enzalutamide complex

3.5 In-silico evaluation of ODM-201 – AR mutant responses

To test the robustness and applicability of the developed methodology, the models were tested against an independent dataset. This section of the study aims at predicting the effect of ODM-201 on the previously reported panel of AR mutants in order to classify the drug-mutant responses as agonist or antagonist. ODM-201 is an anti-androgen that targets the AR pathway, currently under clinical trials.¹⁶³

The panel of 24 mutants along with the WT was docked with ODM-201. A methodology, similar to the pipeline described in Figure 2.2 was employed for this study. The accuracy of the QSAR models for the above described dataset was 72% (73% upon addition of mutant T878G to the test dataset). Importantly, ODM-201 was not used in model training. The computational predictions were established before the availability of experimental results, that illustrates the high predicting power of our approach. Analysis of a blinded independent test represents the applicability of the methodology and unbiased predictions.

Chapter 4: T878G mutant agonizes anti-androgens

The T878G mutant was predicted to yield agonist response to the anti-androgens: Bicalutamide, Enzalutamide and Hydroxyflutamide.⁹ (Figure 4.1)

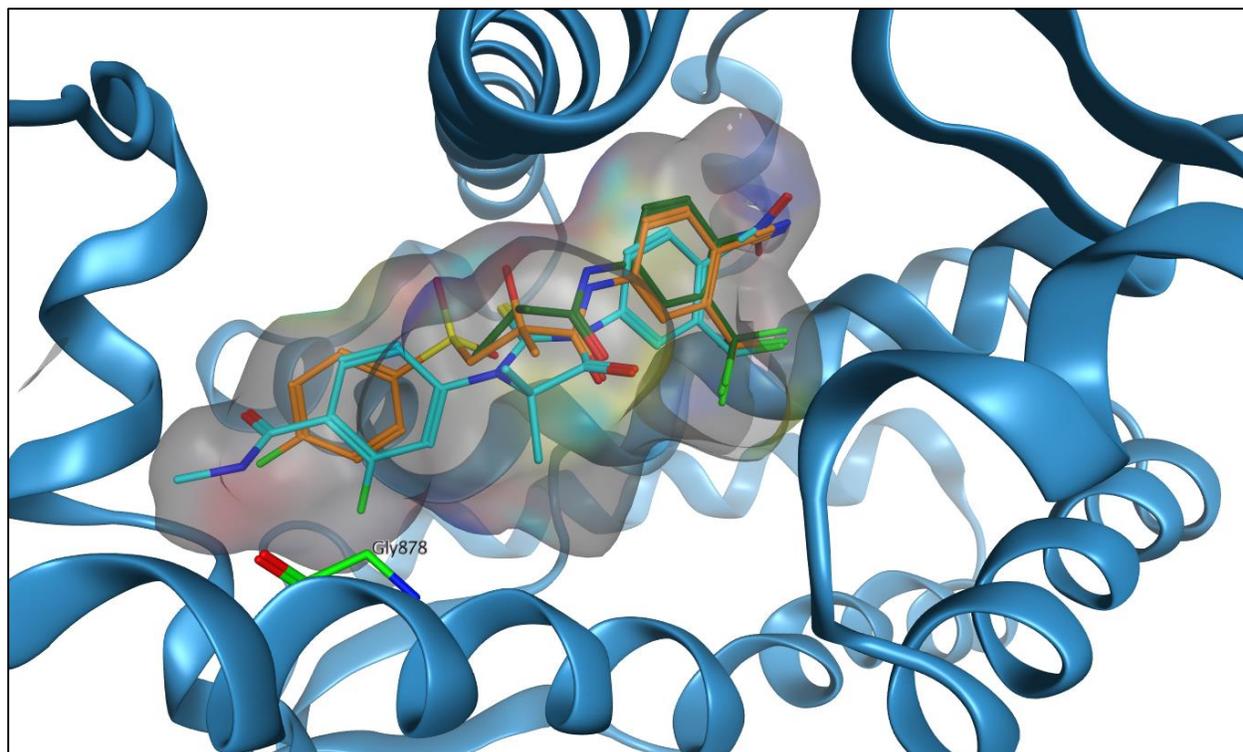


Figure 4.1: Bicalutamide, Enzalutamide and Hydroxyflutamide bound to ABS of T878G mutant

4.1 Structural analysis of T878G – ligand interactions

With no prior information or knowledge about the T878G mutant, one critical fact to be established was the biological viability of this mutant, predicted to be stable and biologically active through in-silico modeling. Therefore, interaction between T878G and DHT was explored. Molecular docking revealed that the DHT binds well within the hydrophobic ABS pocket, with a Glide XP docking score of -8.92 kcal/mol. The structural model shows that DHT binds in its native agonist conformation as seen in most of the AR mutants as well as the wild-type with the backbone of L874 interacting with the hydroxyl group. (See Figure 4.2)

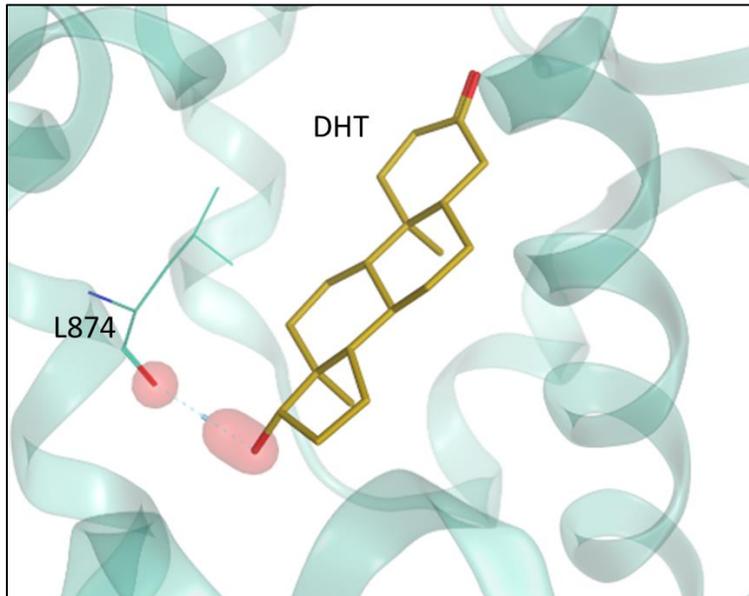


Figure 4.2: T878G mutant activated by DHT, H-bond formation with L874 residue

In terms of anti-androgen binding, Bicalutamide was found to attain a binding conformation similar to the T878G-DHT modeled structure. Despite of a higher molecular weight and greater molecular structural complexity compared to DHT, it further slides down into the ABS pocket; the docking score obtained was -12.10 kcal/mol. The two benzene rings of Bicalutamide stack over the benzene rings of residues F765, F892 and W742. The resultant π - π interactions can potentially explain the sustenance of agonist binding conformation of Bicalutamide within the ABS pocket, in addition to the sidechain H-bond interaction with N706. (See Figure 4.3)

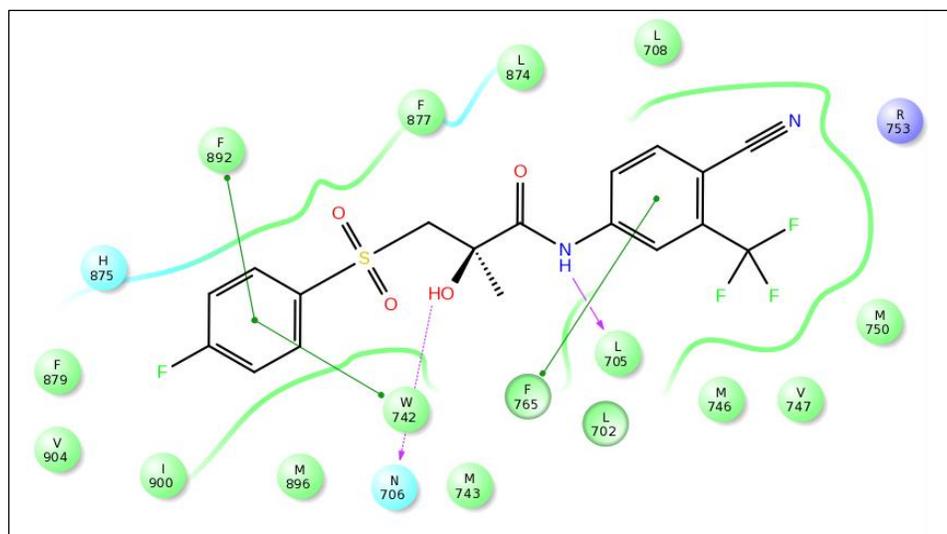


Figure 4.3: Bicalutamide interacts with the T878G ABS pocket through π - π stacking interactions with benzene rings of F892, W742, F765 and H-Bond interaction with N706 sidechain and backbone of L705

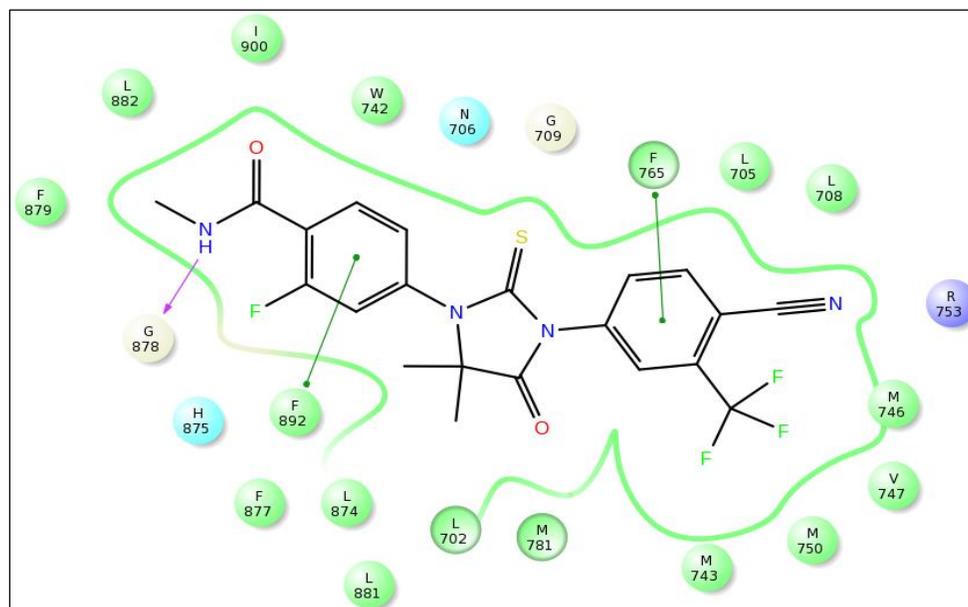


Figure 4.4: Enzalutamide interacts with the T878G mutant pocket by H-bond formation with G878 and π - π stacking interactions between the benzene rings of F765 and F892

On the other hand, Enzalutamide interacts with the T878G mutant by Hydrogen bond formation with the backbone of G878 residue. The residue 878 (T in the Wild-type) is not involved in any kind of direct interaction with the WT receptor. This can be linked to the predicted and experimentally observed agonist response, distinct from the WT. Similar to Bicalutamide, π - π stacking interactions with F765 and F892 potentially stabilize the agonist binding conformation of Enzalutamide aligned inside the ABS pocket. (See Figure 4.5) Apart from the aforementioned interactions, the H-bond formation between G878 and amide moiety was unique to this mutant. The residue 878 in the WT receptor does not interact with Enzalutamide unlike the T878G receptor. (See Figure 4.4) Also, Enzalutamide binds into the ABS in an agonist conformation that indicates towards agonist response characterization.

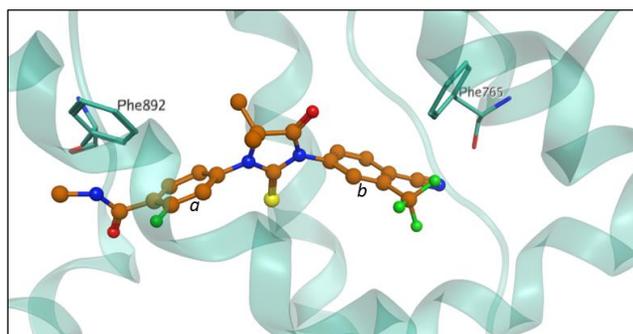


Figure 4.5: Enzalutamide bound to T878G ABS, benzene rings of F892 and F765 interact through T-shaped π - π stacking with *a* and *b* rings of Enzalutamide

Hydroxyflutamide weighs about 292.21 g/mol very similar to the native ABS pocket ligand, DHT (290.44 g/mol). The size and weight of this anti-androgen can be one of the reasons it aligns itself in an agonist-like conformation within the T878G- ABS pocket. A π - π stacking

interaction with the benzene ring of F765 residue was seen with the phenyl propanamide moiety. (See Figure 4.6) An agonist response was predicted for Hydroxyflutamide-T878G interaction.

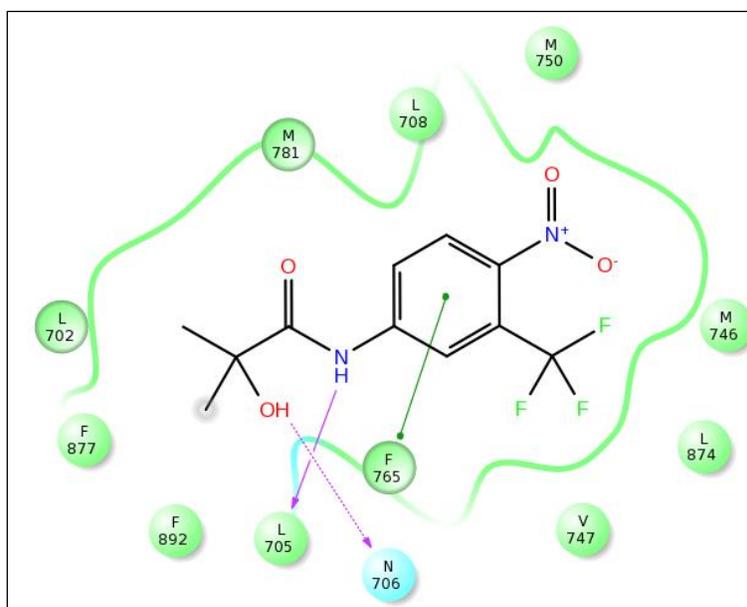


Figure 4.6: Hydroxyflutamide interacting with the T878G mutant ABS pocket through H-bonding with L705 backbone and sidechain of N706, also is seen a π - π stacking interaction with benzene ring of F765

4.2 Descriptor value correlation of T878G – anti-androgen complexes

As observed in case of H875Y mutant descriptor values complexed with different anti-androgens, the T878G mutant was no exception. Hydroxyflutamide poses least amount of steric hindrance, allowing greater mobility within the pocket in contrast to Enzalutamide. As shown in Figure 4.4, Enzalutamide interacts via H-bond formation with the G878 residue, a larger amount of steric interaction value can be seen between them. A biological behavioral scenario similar to the H875Y mutant can be expected, where Hydroxyflutamide acts as a complete agonist failing to inhibit cell growth. (Table 4.1)

T878G	<i>Rs_L_R</i>	<i>Abs_Rs_L_R</i>	<i>878 Rs_AA_L</i>	<i>R_Csp3-L_Fsp3</i>	<i>R_Nsp2-L-Nsp</i>	<i>Predicted response</i>
<i>Bicalutamide</i>	208.26	61.83	0.39	11.86	1.07	Agonist
<i>Enzalutamide</i>	231.75	68.67	0.65	11.76	1.03	Agonist
<i>Hydroxyflutamide</i>	139.23	41.47	0.21	8.93	0.00	Agonist

Table 4.1: Descriptor value correlation with predicted biological response for mutant T878G

4.3 Predicted and experimental biological responses

The mutant T878G was biologically stable and viable. Consensus voting results among the QSAR model predictions were ascertained as predicted activities for the mutant-ligand complexes. The predicted responses were then experimentally tested. The experimental biological activity aligns completely with the predicted response, as shown in Table 4.2.

<i>S. No.</i>	<i>Molecule</i>	<i>Predicted activity</i>	<i>Experimental activity</i>
1.	DHT	+1	+1
2.	Hydroxyflutamide	+1	+1
3.	Bicalutamide	+1	+1
4.	Enzalutamide	+1	+1
5.	ODM-201	-1	-1

Table 4.2: Predicted and experimental biological responses of T878G mutant to DHT and anti-androgens

4.4 MD simulations analysis of T878G – anti-androgen complexes

Given that no prior information about the mutant T878G was available, hence, molecular dynamics simulations aided the better understanding of the structural parameters that are involved in the biological response characterization.

4.4.1 RMSD analysis determines structural stability

The structural stability was evaluated through RMSD analysis of the MD trajectory generated after 25 ns simulation which was compared to the initial docking conformation of the receptor and the ligand. RMSD analysis of the T878G-Enzalutamide complex confirmed no significant changes in the initial binding conformation. The equilibrated structures were consistent with the initially obtained docked poses, indicated through low RMSD fluctuations.

<i>S.No.</i>	<i>Complex</i>	<i>Backbone (Å)</i>	<i>Ligand (Å)</i>
1.	T878G-Enzalutamide	1.11	1.90
2.	T878G-Hydroxyflutamide	1.30	1.58
3.	T878G-Bicalutamide	1.05	1.34
4.	T878G-DHT	1.15	1.24

Table 4.3: RMSD comparison between initial docked and energy equilibrated structures

Enzalutamide binds within the T878G mutant ABS pocket in an agonist-like conformation, the average RMSD was calculated as 1.90 Å for the 25 ns equilibration run. (See Table 4.3) The trajectories obtained for the T878G backbone and Enzalutamide show that there were no major alterations in the structure when subjected to ambient conditions of temperature and pressure. Furthermore, apart from the consistency with the docked structure, Enzalutamide remained bound to the ABS of the mutant pocket.

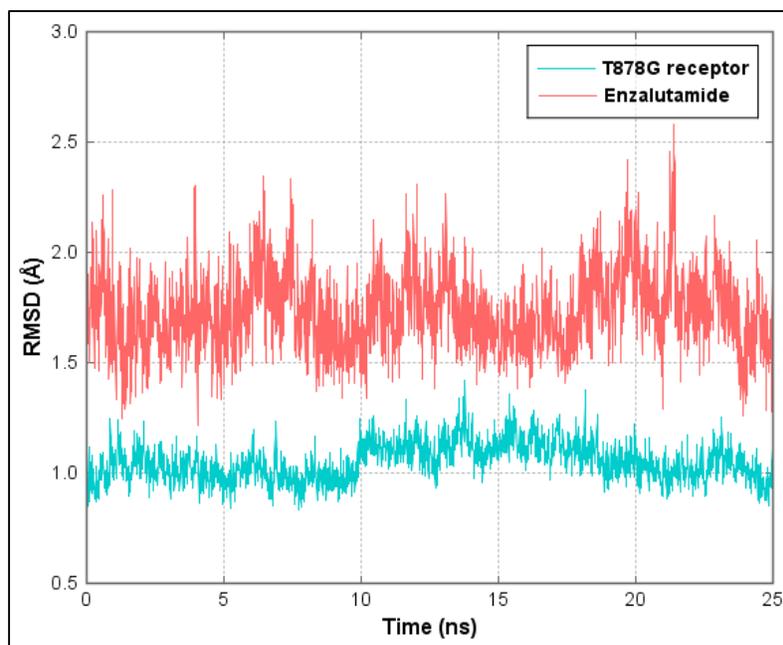


Figure 4.7: Stable trajectory obtained for T878G receptor - Enzalutamide complex

Similar trend could be seen in case of the T878G-Bicalutamide complex where minimal transformations between the initial and final MD structures were observed, represented through stable trajectory. (See Figure 4.8) However, fluctuations were seen in case of the Bicalutamide molecule, which correspond to the flexibility and induced fit effect. The change in RMSD was consistently low 1.34 Å for the Bicalutamide molecule and 1.05 Å for the T878G receptor backbone. As compared to the initial structure, the fluctuation in the equilibrated structures may be due to induced fit effect or flexibility of the protein structure; but nothing major can be concluded since this difference is not major.

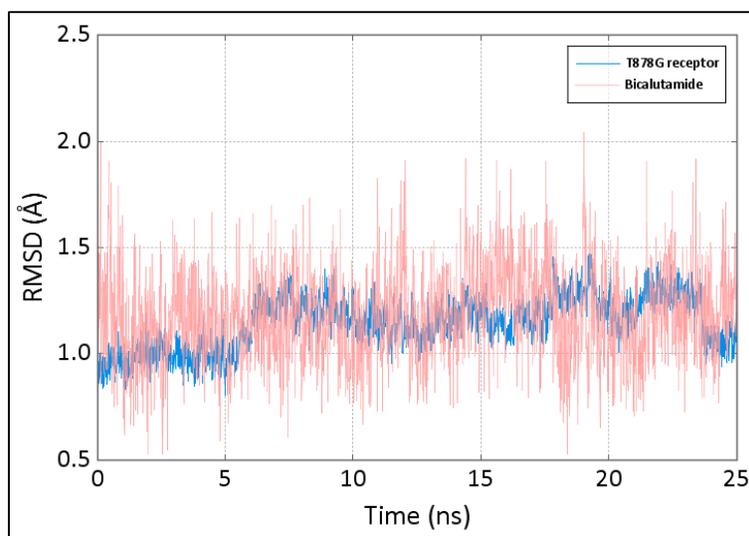


Figure 4.8: RMSD analysis of T878G-Bicalutamide complex

However, in the T878G-Hydroxyflutamide complex, with the ligand molecule being smaller in size compared to Bicalutamide and Enzalutamide, greater fluctuations in the resultant RMSD were observed in both the receptor and the ligand molecule. Hydroxyflutamide did not escape the ABS pocket stabilized within the binding site, after initial conformational changes observed around 12.5 ns down the 25 ns production run. (Figure 4.9) The RMSD measured between the initial docked and equilibrated structure of the T878G receptor was 1.30 Å and 1.58 Å for Hydroxyflutamide, respectively. The receptor structure is highly flexible with the presence of loops, which may be contributing to the overall change in the RMSD. However upon evaluating the ABS pocket of the T878G mutant bound to Hydroxyflutamide no major alterations were observed in the MD trajectory.

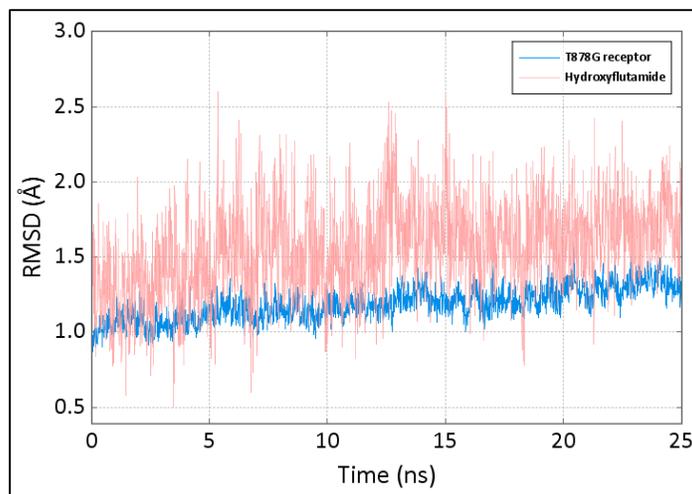


Figure 4.9: RMSD analysis of T878G-Hydroxyflutamide complex

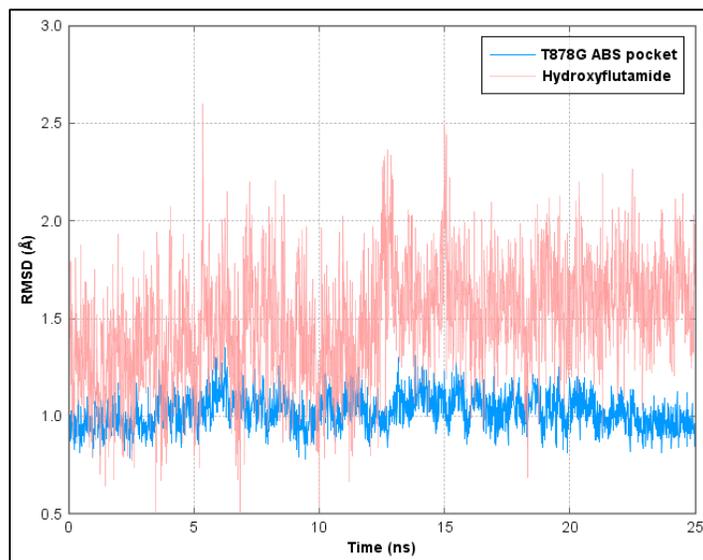


Figure 4.10: No major alterations seen in the MD trajectory of Hydroxyflutamide bound T87G mutant ABS pocket

4.4.2 Contact frequency analysis of T878G – anti-androgen complexes

The contact frequency analysis of the T878G-Enzalutamide complex supports the initial contacts that were obtained after molecular docking. Consistent contacts were observed with residues L705 (100%), N706 (100%), L708 (100%), F765 (99.80%), G878 (99.32%), F892 (97.72%) and

M896 (93.08%). This not only validates the docking pose, but also indicates the key amino acid residues which are critically important for ligand binding and structural integrity of the receptor. Upon modification of key residues such as the F765 or L705, the receptors were structurally denatured, and were not biologically viable. Certain, new contacts were also observed during the MD simulation run, to obtain more conclusive results longer production runs would be required. (Figure 4.11)

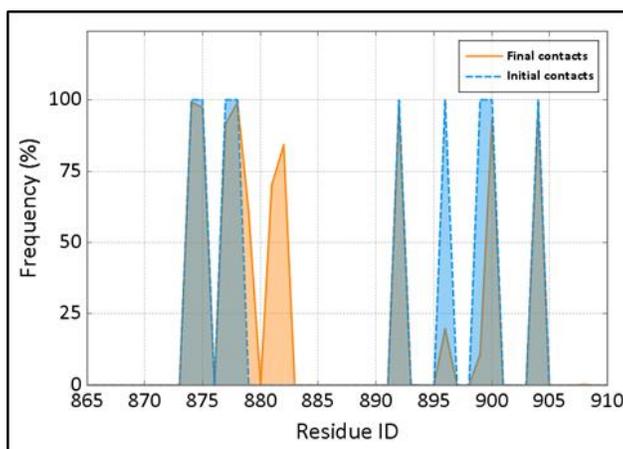


Figure 4.11: Contact frequency comparison for T878G-Enzalutamide complex

4.5 Experimental evaluation of the T878G mutant response to anti-androgens

In order to validate the *in silico* predictions, the response of T878G mutant to anti-androgens was evaluated experimentally using a luciferase reporter transcription assay. PC3 human prostate cancer cells lacking the endogenous AR were transiently transfected with either wild-type or T878G mutated AR.⁹ Cells were stimulated with the non-metabolizable androgen R1881 and then treated with the increasing concentrations of Enzalutamide, Hydroxyflutamide or Bicalutamide. The activation of the T878G mutant was normalized to the WT shown in (%) along the y-axis. (See Figure 4.12) The concentrations used to study the inhibition range between 0 and 50 μ M, shown on x-axis. For Enzalutamide (Figure 4.12 A), an initial inhibition of T878G

was observed however, at higher concentrations (16 μM), the receptor was activated by the anti-androgen, eliciting an agonist response. Hydroxyflutamide on the other hand, yielded an agonist response even at concentrations as low as $7.6 \times 10^{-3} \mu\text{M}$ only to activate the receptor up to four times (400%) as compared to the WT at 50 μM concentration. (Figure 4.12 B) Similarly, Bicalutamide showed initial activation of the T878G receptor at low concentrations gradually inhibiting the growth but activating the receptor at higher concentrations (16.66 – 50 μM range). (See Figure 4.12 C) As predicted T878G mutant showed an agonist response in presence of the 3 evaluated drugs.

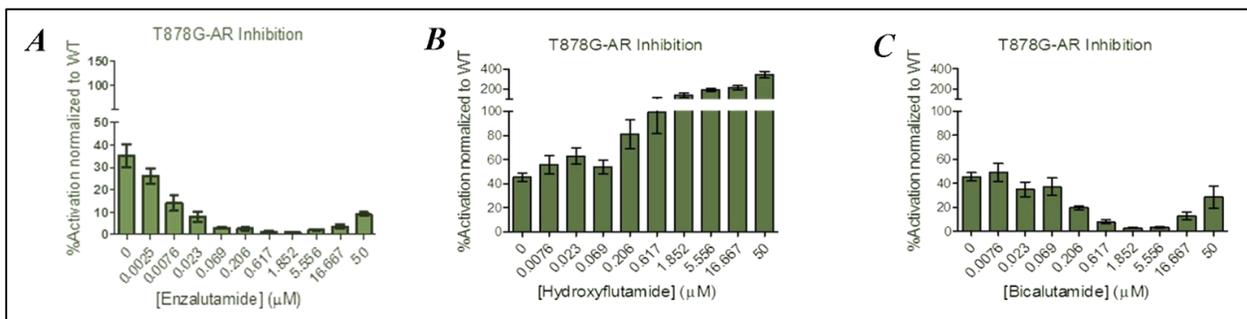


Figure 4.12: The response of T878G mutant to Enzalutamide, Hydroxyflutamide, and Bicalutamide, in an in vitro cell-based assay. Each concentration was assayed in quadruplicate $n = 4$, with a biological replicate of $n = 2$.

4.6 ODM-201 is 'effective' against the T878G mutant

ODM-201 bound distinctively to T878G receptor as compared to other anti-androgens. With no model available, to compare the binding of the ligand molecule, MD simulations were performed to evaluate the predicted docking pose. A stable trajectory was obtained for the T878G pocket in the T878G-ODM complex, with RMSD measured as 0.63 Å. However, due to the presence of various rotatable bonds, and moieties in the ODM-201 structure, the molecular flexibility was taken into account and the RMSD was measured as 1.99 Å between the docked and equilibrated structure.

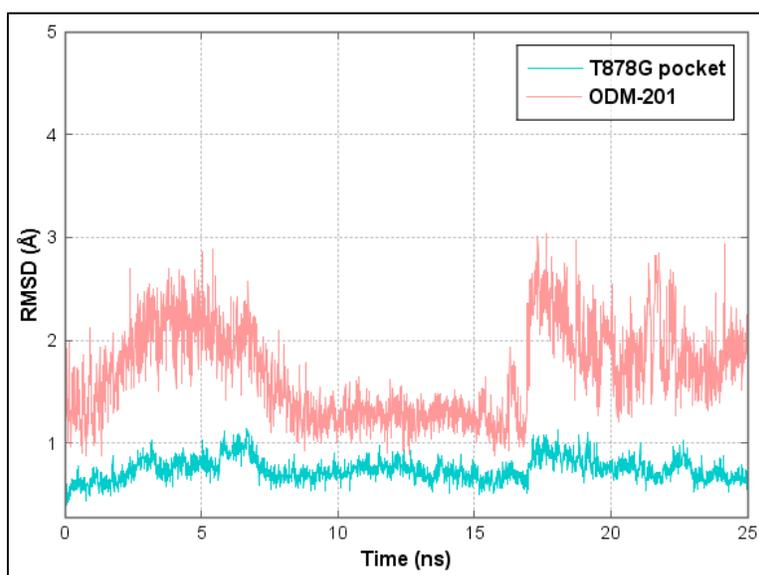


Figure 4.13: Stable trajectory of the T878G pocket bound to ODM-201

ODM-201 was predicted to yield the anticipated antagonist response. The T878G receptor upon stimulation with a non-metabolizable androgen R1881 showed no initial activation by ODM-201. With slight increase in the concentration ($7.6 \times 10^{-3} \mu\text{M}$), ODM-201 yielded the predicted response of antagonism, inhibiting the growth of the T878G mutant. Complete inhibition was achieved at a concentration as low as $1.85 \mu\text{M}$. Thus, the experimental testing confirmed the

predicted response, inhibiting the cell growth in case of T878G mutant upon treatment with ODM-201.)

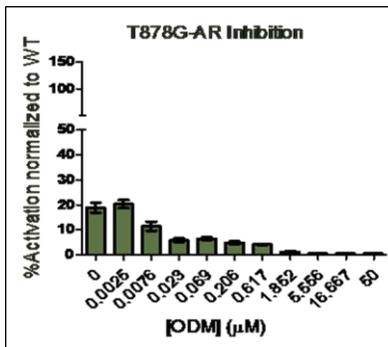


Figure 4.14: ODM-201 yields antagonist response towards T878G mutant

Chapter 5: Conclusions

5.1 Summary of the study

The importance of a prognostic platform for the treatment management of CRPC patients cannot be underestimated. PCa patients can suffer due to inadequacy of treatment response information. With the help of the developed pipeline the prediction and experimental characterization of therapeutic response to various AR mutants was accomplished. The amalgamation of cheminformatics approach with biological testing provided higher speed and efficiency in determining novel AR mutants, which otherwise would be a tedious experimental exercise to create all mutants and test their behavior.

The evidence-based methodology developed through this study could potentially help identifying distinct AR-LBD mutants drug response profiles. Identification of mutants can broaden the ligand specificity of novel treatment options being developed that target the AR pathway for PCa treatment. The resistance to currently used anti-androgens results in adverse clinical effects. AR mutants have been reported to exhibit enhanced activation by anti-androgens accompanied by elevated sensitivity to DHT stimulation as well as to other steroids. This study accomplished the identification of a previously uncharacterized mutant T878G, computational prediction and experimental verification of the drug responses. Patients need to be monitored for this mutant, and suggestive recommendations could be made depending upon the obtained behavioral drug profiles. This study was extended to characterizing therapeutic responses to experimental anti-androgens such as ODM-201, whose response to the panel of AR mutants was unexplored at the time of in-silico analyses. Rigorous exploratory data analysis guides the pipeline to generate therapeutic response predictions for anti-androgens that were not a part of the training dataset. This opens up an avenue for characterizing the responses to other chemical agents and molecules

that target the ABS to achieve AR inhibition for prostate cancer treatment. Furthermore, this approach helps to bypass the time consuming experimental testing of all the possible mutant cases, narrowing down to only a handful. This could save both computational and experimental resources in addition to faster, more accurate mutant — drug response characterization.

5.2 Novel mutants predicted

The characterization T878G mutant's biological response improves our understanding of how the substitution of one amino acid residue by another may entirely alter the biological response to anti-androgens. The substitution of threonine by glycine exposes the amino acid backbone, resulting in elevated levels of protein — ligand interactions. This can be ascribed to the biological response predicted and experimentally validated.

Interestingly, if the T878G mutant appears alongside a previous mutation such as the H875Y; a strong agonist response has been predicted by the QSAR models. This apparently would be the case since the H875Y mutated AR yields strong agonist responses to all 3 anti-androgens. With another mutated residue within the AR, we would expect the response to be stronger (complete agonist, in this case) than a single point mutation in the AR. Anti-androgens have been shown to ineffective at large against the dual point mutants such as the F877L/T878A, H875Y/T878A etc. Furthermore, this pipeline could also predict that the experimental anti-androgen ODM-201 is effective generating antagonist response against the T878G mutant. This prediction was experimentally validated.

A prognostic approach could therefore provide recommendations for the clinicians aiming towards better, personalized management of CRPC patients. 9 other mutants have also been predicted which remain to be experimentally tested for their therapeutic responses.

5.3 Importance of drug response characterization and future scope

In this study, the pipeline focused towards characterizing the response of unknown mutants towards currently used anti-androgens based on the knowledge extracted from the previously known AR mutant-drug response characterization experiment. It has been successfully implemented to characterize the response another experimental anti-androgen (ODM-201) which was not a part of the training dataset. The in-silico engineering of the mutants as well as response characterization is critical for saving time and experimental resources, as the experimental creation and evaluation of a single mutant takes approximately 4 weeks ideally. This an intensive exercise involving creation of mutant constructs, transfection into cells, stimulation and inhibition characterization. As for this study, each inhibition concentration was assayed in quadruplicate ($n = 4$), with a biological replicate of ($n = 2$). This pipeline can further be extended into various dedicated workflows by addition of more biological end points, continuous variables, anti-androgens, and new mutants etc. which can accelerate and strengthen the potential applicability of the developed methodology.

This approach can also help improve our understanding of broadened ligand specificity to various identified mutations, which could be translated into a foreground for the creation of novel, more potent anti-androgens with no or minimal cross-resistance.

Bibliography

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics.; desLibris - Documents, Canadian Cancer Statistics 2015 - Special Topic Predictions of the Future Burden of Cancer in Canada. Canadian Cancer Society.; S.I., 2015; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC_023395514&title=Canadian%20Cancer%20Statistics%202015%20-%20Special%20Topic%203A%20Predictions%20of%20the%20Future%20Burden%20of%20Cancer%20in%20Canada. (visited on 08/10/2016)
2. Ellison, L., Prostate cancer trends in Canada, 1995 to 2012. In *Health at a Glance*, Statistics Canada 2016.
3. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer Statistics, 2016. *Ca-Cancer J Clin* **2016**, *66* (1), 7-30.
4. Beltran, H.; Beer, T. M.; Carducci, M. A.; de Bono, J.; Gleave, M.; Hussain, M.; Kelly, W. K.; Saad, F.; Sternberg, C.; Tagawa, S. T.; Tannock, I. F., New Therapies for Castration-Resistant Prostate Cancer: Efficacy and Safety. *European urology* **2011**, *60* (2), 279-290.
5. Beltran, H.; Yelensky, R.; Frampton, G. M.; Park, K.; Downing, S. R.; MacDonald, T. Y.; Jarosz, M.; Lipson, D.; Tagawa, S. T.; Nanus, D. M.; Stephens, P. J.; Mosquera, J. M.; Cronin, M. T.; Rubin, M. A., Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. *European urology* **2013**, *63* (5), 920-6.
6. Wong, Y. N.; Ferraldeschi, R.; Attard, G.; de Bono, J., Evolution of androgen receptor targeted therapy for advanced prostate cancer. *Nature reviews. Clinical oncology* **2014**, *11* (6), 365-76.
7. Lorente, D.; Mateo, J.; Zafeiriou, Z.; Smith, A. D.; Sandhu, S.; Ferraldeschi, R.; de Bono, J. S., Switching and withdrawing hormonal agents for castration-resistant prostate cancer. *Nature reviews. Urology* **2015**, *12* (1), 37-47.
8. Lallous, N.; Volik, S. V.; Awrey, S.; Leblanc, E.; Tse, R.; Murillo, J.; Singh, K.; Azad, A. A.; Wyatt, A. W.; LeBihan, S.; Chi, K. N.; Gleave, M. E.; Rennie, P. S.; Collins, C. 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* **2016**, *17*, 10.
9. Paul, N.; Carabet, L. A.; Lallous, N.; Yamazaki, T.; Gleave, M. E.; Rennie, P. S.; Cherkasov, A., Cheminformatics Modeling of Adverse Drug Responses by Clinically Relevant Mutants of Human Androgen Receptor. *Journal of chemical information and modeling* **2016**.
10. Martini, F.; Timmons, M. J.; Tallitsch, R. B., *Human anatomy*. 6th ed.; Pearson Benjamin Cummings: San Francisco, 2009; p xxxiii, 869 p.
11. Shier, D.; Hole, J. W.; Butler, J.; Lewis, R., *Hole's essentials of human anatomy and physiology*. 7th ed.; McGraw-Hill: Boston, Mass, 2000; p xxiv, 613 p.
12. Nickel, J. C., *The prostatitis manual : a practical guide to management of prostatitis/chronic pelvic pain syndrome*. Bladon Medical Pub.: Oxfordshire, UK, 2002; p ix, 116 pages.

13. Mujoomdar, M.; Spry, C.; Canadian Agency for Drugs and Technologies in Health Health Technology Inquiry Service.; desLibris - Documents, Greenlight laser for benign prostatic hypertrophy a clinical and cost-effectiveness review. Canadian Agency for Drugs and Technologies in Health, Health Technology Inquiry Service (HTIS); S.I., 2010; p. 1 online resource.
http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000912038&title=Greenlight%20laser%20for%20benign%20prostatic%20hypertrophy%20a%20clinical%20and%20cost-effectiveness%20review. (visited on 08/11/2016)
14. Grimm, P. D.; Blasko, J. C.; Sylvester, J. E., *The prostate cancer treatment book : advice from leading prostate experts from the nation's top medical institutions*. McGraw Hill: New York, 2004; p xv, 224 p.
15. Thompson, I.; Leach, R. J.; Pollock, B. H.; Naylor, S. L., Prostate cancer and prostate-specific antigen: the more we know, the less we understand. *J Natl Cancer Inst* **2003**, *95* (14), 1027-8.
16. Jones, J. S.; SpringerLink ebooks - Medicine (2013), Prostate cancer diagnosis PSA, biopsy and beyond. In *Current clinical urology* [Online] Humana Press,; New York, 2013; p. 1 online resource.
http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000796957&title=Prostate%20Cancer%20Diagnosis%3A%20PSA%2C%20Biopsy%20and%20Beyond. (visited on 08/11/2016)
17. Pacheco, S. O.; Pacheco, F. J.; Zapata, G. M.; Garcia, J. M.; Previale, C. A.; Cura, H. E.; Craig, W. J., Food Habits, Lifestyle Factors, and Risk of Prostate Cancer in Central Argentina: A Case Control Study Involving Self-Motivated Health Behavior Modifications after Diagnosis. *Nutrients* **2016**, *8* (7).
18. Liss, M. A.; Schenk, J. M.; Faino, A. V.; Newcomb, L. F.; Boyer, H.; Brooks, J. D.; Carroll, P. R.; Dash, A.; Fabrizio, M. D.; Gleave, M. E.; Nelson, P. S.; Neuhaus, M. L.; Wei, J. T.; Zheng, Y.; Wright, J. L.; Lin, D. W.; Thompson, I. M., A diagnosis of prostate cancer and pursuit of active surveillance is not followed by weight loss: potential for a teachable moment. *Prostate cancer and prostatic diseases* **2016**.
19. Eastham, J. A.; Schaeffer, E. M.; SpringerLink (Online service); SpringerLINK ebooks - Medicine (2014), Radical Prostatectomy Surgical Perspectives. Springer New York,; S.I., 2014; p. 1 online resource.
http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001067889&title=Radical%20Prostatectomy%20Surgical%20Perspectives. (visited 08/12/2016)
20. Dunn, M. W.; Kazer, M. W., Prostate cancer overview. *Seminars in oncology nursing* **2011**, *27* (4), 241-50.
21. Carter, H. B.; Isaacs, J. T., Overview of hormonal therapy for prostate cancer. *Progress in clinical and biological research* **1990**, *350*, 129-40.
22. Bishr, M.; Saad, F., Overview of the latest treatments for castration-resistant prostate cancer. *Nature reviews. Urology* **2013**, *10* (9), 522-8.
23. Gao, W.; Bohl, C. E.; Dalton, J. T., Chemistry and structural biology of androgen receptor. *Chem. Rev.* **2005**, *105* (9), 3352-70.
24. Papaioannou, M.; Schleich, S.; Roell, D.; Schubert, U.; Tanner, T.; Claessens, F.; Matusch, R.; Baniahmad, A., NBBS isolated from *Pygeum africanum* bark exhibits androgen

- antagonistic activity, inhibits AR nuclear translocation and prostate cancer cell growth. *Investigational new drugs* **2010**, *28* (6), 729-43.
25. Luke, M. C.; Coffey, D. S., Human androgen receptor binding to the androgen response element of prostate specific antigen. *Journal of andrology* **1994**, *15* (1), 41-51.
26. Crona, D. J.; Milowsky, M. I.; Whang, Y. E., Androgen receptor targeting drugs in castration-resistant prostate cancer and mechanisms of resistance. *Clinical pharmacology and therapeutics* **2015**, *98* (6), 582-9.
27. Tan, M. H.; Li, J.; Xu, H. E.; Melcher, K.; Yong, E. L., Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta pharmacologica Sinica* **2015**, *36* (1), 3-23.
28. Lavery, D. N.; McEwan, I. J., Structural characterization of the native NH₂-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry-Us* **2008**, *47* (11), 3360-9.
29. Monks, D. A.; Rao, P.; Mo, K.; Johansen, J. A.; Lewis, G.; Kemp, M. Q., Androgen receptor and Kennedy disease/spinal bulbar muscular atrophy. *Hormones and behavior* **2008**, *53* (5), 729-40.
30. He, B.; Kempainen, 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* (30), 22986-22994.
31. Schaufele, F.; Carbonell, X.; Guerbardot, M.; Borngraeber, S.; Chapman, M. S.; Ma, A. A.; Miner, J. N.; Diamond, M. I., The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (28), 9802-7.
32. Imamura, Y.; Sadar, M. D., Androgen receptor targeted therapies in castration-resistant prostate cancer: Bench to clinic. *International journal of urology : official journal of the Japanese Urological Association* **2016**, *23* (8), 654-65.
33. Li, H. F.; Ban, F. Q.; 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* (15), 6458-6467.
34. Carsonjurica, M. A.; Schrader, W. T.; Omalley, B. W., Steroid-Receptor Family - Structure and Functions. *Endocr Rev* **1990**, *11* (2), 201-220.
35. Schoenmakers, E.; Alen, P.; Verrijdt, G.; Peeters, B.; Verhoeven, G.; Rombauts, W.; Claessens, F., Differential DNA binding by the androgen and glucocorticoid receptors involves the second Zn-finger and a C-terminal extension of the DNA-binding domains. *Biochem J* **1999**, *341*, 515-521.
36. Claessens, F.; Denayer, S.; Van Tilborgh, N.; Kerkhofs, S.; Helsen, C.; Haelens, A., Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nuclear receptor signaling* **2008**, *6*, e008.
37. Bevan, C. L.; Hoare, S.; Claessens, F.; Heery, D. M.; Parker, M. G., The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. *Mol Cell Biol* **1999**, *19* (12), 8383-92.
38. Srinivas-Shankar, U.; Wu, F. C. W., Drug insight: testosterone preparations. *Nature Clinical Practice Urology* **2006**, *3* (12), 653-665.

39. Shang, Y. F.; Myers, M.; Brown, M., Formation of the androgen receptor transcription complex. *Mol Cell* **2002**, *9* (3), 601-610.
40. 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* (17), 6201-6206.
41. De Jesus-Tran, K. P.; 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 Sci* **2006**, *15* (5), 987-999.
42. Jaaskelainen, J.; Deeb, A.; Schwabe, J. W.; Mongan, N. P.; Martin, H.; Hughes, I. A., Human androgen receptor gene ligand-binding-domain mutations leading to disrupted interaction between the N- and C-terminal domains. *J Mol Endocrinol* **2006**, *36* (2), 361-8.
43. Hay, C. W.; McEwan, I. J., The impact of point mutations in the human androgen receptor: classification of mutations on the basis of transcriptional activity. *PloS one* **2012**, *7* (3), e32514.
44. Nguyen, P. L.; Alibhai, S. M. H.; Basaria, S.; D'Amico, A. V.; Kantoff, P. W.; Keating, N. L.; Penson, D. F.; Rosario, D. J.; Tombal, B.; Smith, M. R., Adverse Effects of Androgen Deprivation Therapy and Strategies to Mitigate Them. *European urology* **2015**, *67* (5), 825-836.
45. Botrel, T. E.; Clark, O.; Lima Pompeo, A. C.; Horta Bretas, F. F.; Sadi, M. V.; Ferreira, U.; Borges Dos Reis, R., Efficacy and Safety of Combined Androgen Deprivation Therapy (ADT) and Docetaxel Compared with ADT Alone for Metastatic Hormone-Naive Prostate Cancer: A Systematic Review and Meta-Analysis. *PloS one* **2016**, *11* (6), e0157660.
46. Muralidhar, V.; Regan, M. M.; Werner, L.; Nakabayashi, M.; Evan, C. P.; Bellmunt, J.; Choueiri, T. K.; Elfiky, A. A.; Harshman, L. C.; McKay, R. R.; Pomerantz, M. M.; Sweeney, C. J.; Taplin, M. E.; Kantoff, P. W.; Nguyen, P. L., Duration of Androgen Deprivation Therapy for High-Risk Prostate Cancer: Application of Randomized Trial Data in a Tertiary Referral Cancer Center. *Clinical genitourinary cancer* **2016**, *14* (4), e299-305.
47. Grasso, C. S.; Wu, Y. M.; Robinson, D. R.; Cao, X.; Dhanasekaran, S. M.; Khan, A. P.; Quist, M. J.; Jing, X.; Lonigro, R. J.; Brenner, J. C.; Asangani, I. A.; Ateeq, B.; Chun, S. Y.; Siddiqui, J.; Sam, L.; Anstett, M.; Mehra, R.; Prensner, J. R.; Palanisamy, N.; Ryslik, G. A.; Vandin, F.; Raphael, B. J.; Kunju, L. P.; Rhodes, D. R.; Pienta, K. J.; Chinnaiyan, A. M.; Tomlins, S. A., The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **2012**, *487* (7406), 239-43.
48. Bazarbashi, S.; Bachour, M.; Bulbul, M.; Alotaibi, M.; Jaloudi, M.; Jaafar, H.; Mukherji, D.; Farah, N.; Alrubai, T.; Shamseddine, A., Metastatic castration resistant prostate cancer: current strategies of management in the Middle East. *Critical reviews in oncology/hematology* **2014**, *90* (1), 36-48.
49. Drake, C. G.; Sharma, P.; Gerritsen, W., Metastatic castration-resistant prostate cancer: new therapies, novel combination strategies and implications for immunotherapy. *Oncogene* **2014**, *33* (43), 5053-64.
50. Altavilla, A.; Iacovelli, R.; Procopio, G.; Alesini, D.; Risi, E.; Campenni, G. M.; Palazzo, A.; Cortesi, E., Medical strategies for treatment of castration resistant prostate cancer (CRPC) docetaxel resistant. *Cancer biology & therapy* **2012**, *13* (11), 1001-8.

51. Mansi, L.; Thiery-Vuillemin, A.; Kalbacher, E.; Nguyen, T.; Maurina, T.; Nallet, J.; Kim, S.; Borg, C.; Kleinclauss, F.; Pivot, X.; Adotevi, O., [Immunotherapy: an emerging strategies against prostate castration resistant cancer]. *Bulletin du cancer* **2012**, *99 Suppl 1*, S57-65.
52. Fujimoto, N.; Shiota, M.; Kubo, T.; Matsumoto, T., Novel therapeutic strategies following docetaxel-based chemotherapy in castration-resistant prostate cancer. *Expert review of clinical pharmacology* **2010**, *3* (6), 785-95.
53. Tanaka, T.; Nakatani, T., New therapeutic strategies for castration-resistant prostate cancer. *Recent patents on anti-cancer drug discovery* **2011**, *6* (3), 373-83.
54. Wang, H. T.; Yao, Y. H.; Li, B. G.; Tang, Y.; Chang, J. W.; Zhang, J., Neuroendocrine Prostate Cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis-a systematic review and pooled analysis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2014**, *32* (30), 3383-90.
55. Robinson, D.; Van Allen, E. M.; Wu, Y. M.; Schultz, N.; Lonigro, R. J.; Mosquera, J. M.; Montgomery, B.; Taplin, M. E.; Pritchard, C. C.; Attard, G.; Beltran, H.; Abida, W.; Bradley, R. K.; Vinson, J.; Cao, X.; Vats, P.; Kunju, L. P.; Hussain, M.; Feng, F. Y.; Tomlins, S. A.; Cooney, K. A.; Smith, D. C.; Brennan, C.; Siddiqui, J.; Mehra, R.; Chen, Y.; Rathkopf, D. E.; Morris, M. J.; Solomon, S. B.; Durack, J. C.; Reuter, V. E.; Gopalan, A.; Gao, J.; Loda, M.; Lis, R. T.; Bowden, M.; Balk, S. P.; Gaviola, G.; Sougnez, C.; Gupta, M.; Yu, E. Y.; Mostaghel, E. A.; Cheng, H. H.; Mulcahy, H.; True, L. D.; Plymate, S. R.; Dvinge, H.; Ferraldeschi, R.; Flohr, P.; Miranda, S.; Zafeiriou, Z.; Tunariu, N.; Mateo, J.; Perez-Lopez, R.; Demichelis, F.; Robinson, B. D.; Schiffman, M.; Nanus, D. M.; Tagawa, S. T.; Sigaras, A.; Eng, K. W.; Elemento, O.; Sboner, A.; Heath, E. I.; Scher, H. I.; Pienta, K. J.; Kantoff, P.; de Bono, J. S.; Rubin, M. A.; Nelson, P. S.; Garraway, L. A.; Sawyers, C. L.; Chinnaiyan, A. M., Integrative clinical genomics of advanced prostate cancer. *Cell* **2015**, *161* (5), 1215-28.
56. Waltering, K. K.; Urbanucci, A.; Visakorpi, T., Androgen receptor (AR) aberrations in castration-resistant prostate cancer. *Molecular and cellular endocrinology* **2012**, *360* (1-2), 38-43.
57. Chen, Z.; Lan, X.; Thomas-Ahner, J. M.; Wu, D. Y.; Liu, X. T.; Ye, Z. Q.; Wang, L. G.; Sunkel, B.; Grenade, C.; Chen, J. S.; Zynger, D. L.; Yan, P. S.; Huang, J. T.; Nephew, K. P.; Huang, T. H. M.; Lin, S. L.; Clinton, S. K.; Li, W.; Jin, V. X.; Wang, Q. B., Agonist and antagonist switch DNA motifs recognized by human androgen receptor in prostate cancer. *Embo J* **2015**, *34* (4), 502-516.
58. Culig, Z.; Hoffmann, J.; Erdel, M.; Eder, I. E.; Hobisch, A.; Hittmair, A.; Bartsch, G.; Utermann, G.; Schneider, M. R.; Parczyk, K.; Klocker, H., Switch from antagonist to agonist of the androgen receptor blocker bicalutamide is associated with prostate tumour progression in a new model system. *Brit J Cancer* **1999**, *81* (2), 242-251.
59. Eisermann, K.; Wang, D.; Jing, Y.; Pascal, L. E.; Wang, Z., Androgen receptor gene mutation, rearrangement, polymorphism. *Translational andrology and urology* **2013**, *2* (3), 137-147.
60. Wyatt, A. W.; Azad, A. A.; Volik, S. V.; Annala, M.; Beja, K.; McConeghy, B.; Haegert, A.; Warner, E. W.; Mo, F.; Brahmhatt, S.; Shukin, R.; Le Bihan, S.; Gleave, M. E.; Nykter, M.; Collins, C. C.; Chi, K. N., Genomic Alterations in Cell-Free DNA and Enzalutamide Resistance in Castration-Resistant Prostate Cancer. *JAMA oncology* **2016**.

61. Emes, R. D.; Riley, M. C.; Laukaitis, C. M.; Goodstadt, L.; Karn, R. C.; Ponting, C. P., Comparative evolutionary genomics of androgen-binding protein genes. *Genome research* **2004**, *14* (8), 1516-1529.
62. Korpai, 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* (9), 1030-43.
63. Prekovic, S.; van Royen, M. E.; Voet, A. R. D.; Geverts, B.; Houtman, R.; Melchers, D.; Zhang, K. Y. J.; Van den Broeck, T.; Smeets, E.; Spans, L.; Houtsmuller, A. B.; Joniau, S.; Claessens, F.; Helsen, C., The Effect of F877L and T878A Mutations on Androgen Receptor Response to Enzalutamide. *Molecular cancer therapeutics* **2016**, *15* (7), 1702-1712.
64. Macalino, S. J.; Gosu, V.; Hong, S.; Choi, S., Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research* **2015**, *38* (9), 1686-701.
65. 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* (1), 127-41.
66. Baig, M. H.; Ahmad, K.; Roy, S.; Ashraf, J. M.; Adil, M.; Siddiqui, M. H.; Khan, S.; Kamal, M. A.; Provaznik, I.; Choi, I., Computer Aided Drug Design: Success and Limitations. *Current pharmaceutical design* **2016**, *22* (5), 572-81.
67. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr., Computational methods in drug discovery. *Pharmacological reviews* **2014**, *66* (1), 334-95.
68. Laurie, A. T.; Jackson, R. M., Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Current protein & peptide science* **2006**, *7* (5), 395-406.
69. 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* (7), 1757-68.
70. Vyas, V. K.; Ghate, M.; Patel, K.; Qureshi, G.; Shah, S., Homology modeling, binding site identification and docking study of human angiotensin II type I (Ang II-AT1) receptor. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **2015**, *74*, 42-8.
71. Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M., Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* **2011**, *7* (2), 146-57.
72. Yuriev, E.; Holien, J.; Ramsland, P. A., Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *J Mol Recognit* **2015**, *28* (10), 581-604.
73. Rosenfeld, R.; Vajda, S.; Delisi, C., Flexible Docking and Design. *Annu Rev Bioph Biom* **1995**, *24*, 677-700.
74. Krol, M.; Tournier, A. L.; Bates, P. A., Flexible relaxation of rigid-body docking solutions. *Proteins* **2007**, *68* (1), 159-69.
75. Tinberg, C. E.; Khare, S. D.; Dou, J. Y.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D., Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **2013**, *501* (7466), 212-+.
76. 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* (7), 1739-1749.

77. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry* **2004**, *47* (7), 1750-1759.
78. Goodsell, D. S.; Morris, G. M.; Olson, A. J., Automated docking of flexible ligands: Applications of AutoDock. *J Mol Recognit* **1996**, *9* (1), 1-5.
79. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P., eHITS: An innovative approach to the docking and scoring function problems. *Current protein & peptide science* **2006**, *7* (5), 421-435.
80. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins-Structure Function and Genetics* **2003**, *52* (4), 609-623.
81. Abagyan, R.; Totrov, M.; Kuznetsov, D., Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *Journal of computational chemistry* **1994**, *15* (5), 488-506.
82. Jorgensen, W. L., The many roles of computation in drug discovery. *Science* **2004**, *303* (5665), 1813-1818.
83. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A., Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of medicinal chemistry* **2016**, *59* (9), 4035-61.
84. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC biology* **2011**, *9*, 71.
85. Leach, A. R., *Molecular modelling : principles and applications*. Longman: Harlow, England, 1996; p xvi, 595 p.
86. Carlson, H. A., Protein flexibility and drug design: how to hit a moving target. *Current opinion in chemical biology* **2002**, *6* (4), 447-52.
87. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How Fast-Folding Proteins Fold. *Science* **2011**, *334* (6055), 517-520.
88. Shan, Y. B.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., How Does a Drug Molecule Find Its Target Binding Site? *J Am Chem Soc* **2011**, *133* (24), 9181-9183.
89. Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A., Target flexibility: an emerging consideration in drug discovery and design. *Journal of medicinal chemistry* **2008**, *51* (20), 6237-55.
90. Sukumar, N.; Das, S., Current Trends in Virtual High Throughput Screening Using Ligand-Based and Structure-Based Methods. *Combinatorial chemistry & high throughput screening* **2011**, *14* (10), 872-888.
91. 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. H.; 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* (12), 4977-5010.
92. Hansch, C., Computerized Approach to Quantitative Biochemical Structure-Activity Relationships. *Abstr Pap Am Chem S* **1971**, (Mar-A), 50-&.
93. Hansch, C., A Quantitative Approach to Biochemical Structure-Activity Relationships. *Accounts of chemical research* **1969**, *2* (8), 232-&.

94. Sahlin, U.; Filipsson, M.; Oberg, T., A Risk Assessment Perspective of Current Practice in Characterizing Uncertainties in QSAR Regression Predictions. *Molecular informatics* **2011**, *30* (6-7), 551-64.
95. Pirhadi, S.; Shiri, F.; Ghasemi, J. B., Multivariate statistical analysis methods in QSAR. *Rsc Adv* **2015**, *5* (127), 104635-104665.
96. Tropsha, A., Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular informatics* **2010**, *29* (6-7), 476-488.
97. Gramatica, P., Principles of QSAR models validation: internal and external. *Qsar Comb Sci* **2007**, *26* (5), 694-701.
98. Frimayanti, N.; Yam, M. L.; Lee, H. B.; Othman, R.; Zain, S. M.; Abd Rahman, N., Validation of Quantitative Structure-Activity Relationship (QSAR) Model for Photosensitizer Activity Prediction. *International journal of molecular sciences* **2011**, *12* (12), 8626-8644.
99. Kiralj, R.; Ferreira, M. M. C., Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J Brazil Chem Soc* **2009**, *20* (4), 770-787.
100. Sushko, I.; Novotarskyi, S.; Korner, R.; Pandey, A. K.; Cherkasov, A.; Lo, J. Z.; Gramatica, P.; Hansen, K.; Schroeter, T.; Muller, K. R.; Xi, L. L.; Liu, H. X.; Yao, X. J.; Oberg, T.; Hormozdiari, F.; Dao, P. H.; Sahinalp, C.; Todeschini, R.; Polishchuk, P.; Artemenko, A.; Kuz'min, V.; Martin, T. M.; Young, D. M.; Fourches, D.; Muratov, E.; Tropsha, A.; Baskin, I.; Horvath, D.; Marcou, G.; Muller, C.; Varnek, A.; Prokopenko, V. V.; Tetko, I. V., Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *Journal of chemical information and modeling* **2010**, *50* (12), 2094-2111.
101. Duffy, B. C.; Zhu, L.; Decornez, H.; Kitchen, D. B., Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series. *Bioorganic & medicinal chemistry* **2012**, *20* (18), 5324-5342.
102. Lahana, R., Cheminformatics - decision making in drug discovery. *Drug discovery today* **2002**, *7* (17), 898-900.
103. Mitchell, J. B., Machine learning methods in chemoinformatics. *Wiley interdisciplinary reviews. Computational molecular science* **2014**, *4* (5), 468-481.
104. Lavecchia, A., Machine-learning approaches in drug discovery: methods and applications. *Drug discovery today* **2015**, *20* (3), 318-331.
105. Aha, D. W.; Kibler, D.; Albert, M. K., Instance-Based Learning Algorithms. *Mach Learn* **1991**, *6* (1), 37-66.
106. Breiman, L., Random forests. *Mach Learn* **2001**, *45* (1), 5-32.
107. Chang, C. C.; Lin, C. J., LIBSVM: A Library for Support Vector Machines. *ACM TIST* **2011**, *2* (3).
108. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H., The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* **2009**, *11* (1), 10-18.
109. Iba, W.; Langley, P., Induction of One-Level Decision Trees. *Mach Learn* **1992**, 233-240.
110. Ting, K. M.; Witten, I. H., Stacking Bagged and Dagged Models. In *Proceedings of the Fourteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc.: 1997; pp 367-375.
111. Iba, W.; Langley, P., Induction of One-Level Decision Trees. *Machine Learning* / **1992**, 233-240.

112. Cichosz, P.; Wiley-Blackwell Online Books; Ebrary Academic Complete (Canada) Subscription Collection, Data mining algorithms explained using R. John Wiley & Sons Inc.;; Chichester, West Sussex ; Malden, MA, 2015; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001367765&title=Data%20mining%20algorithms%20%3A%20explained%20using%20R. (visited on 01/15/2016)
113. Sampat, M. P.; Bovik, A. C.; Aggarwal, J. K.; Castleman, K. R., Supervised parametric and non-parametric classification of chromosome images. *Pattern recognition* **2005**, *38* (8), 1209-1223.
114. Spiliopoulou, M.; Schmidt-Thieme, L.; Janning, R.; SpringerLINK ebooks - Mathematics and Statistics (2014), Data Analysis, Machine Learning and Knowledge Discovery. Springer International Publishing,; S.l., 2014; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001067339&title=Data%20Analysis%2C%20Machine%20Learning%20and%20Knowledge%20Discovery. (visited on 01/15/2016)
115. Chang, C. C.; Lin, C. J., LIBSVM: A Library for Support Vector Machines. *Acm T Intel Syst Tec* **2011**, *2* (3).
116. Scholkopf, B., The kernel trick for distances. *Adv Neur In* **2001**, *13*, 301-307.
117. Leach, A. R.; Gillet, V. J.; SpringerLink ebooks - Chemistry and Materials Science (2007); ebrary eBooks, An Introduction to Chemoinformatics. Springer: New York, 2007; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000318465&title=An%20Introduction%20to%20Chemoinformatics. (visited on 12/05/2015)
118. Kurşa, M. B.; Rudnicki, W. R., Feature Selection with the Boruta Package. *J Stat Softw* **2010**, *36* (11), 1-13.
119. Kuhn, M., Building Predictive Models in R Using the caret Package. *J Stat Softw* **2008**, *28* (5), 1-26.
120. Ruffolo, R. R., Important Concepts of Receptor Theory. *J Auton Pharmacol* **1982**, *2* (4), 277-295.
121. Wyatt, A. W.; Gleave, M. E., Targeting the adaptive molecular landscape of castration-resistant prostate cancer. *EMBO molecular medicine* **2015**, *7* (7), 878-894.
122. Wassersug, R. J.; Walker, L. M.; Robinson, J. W.; Ebrary Academic Complete (Canada) Subscription Collection, Androgen deprivation therapy an essential guide for prostate cancer patients and their loved ones. Demos Health,; New York, 2014; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001225984&title=Androgen%20deprivation%20therapy%20%3A%20an%20essential%20guide%20for%20prostate%20cancer%20patients%20and%20their%20loved%20ones. (visited on 06/09/2016)
123. Nelson, D. L.; Cox, M. M.; Lehninger, A. L., *Lehninger principles of biochemistry*. 6th ed.; W.H. Freeman: New York, 2013; p 1198, 101 p.
124. Schwarzenbach, H.; Alix-Panabieres, C.; Muller, I.; Letang, N.; Vendrell, J. P.; Rebillard, X.; Pantel, K., Cell-free Tumor DNA in Blood Plasma As a Marker for Circulating Tumor Cells in Prostate Cancer. *Clinical Cancer Research* **2009**, *15* (3), 1032-1038.
125. Boyd, S., Molecular operating environment. *Chem. World* **2005**, *2* (9), 66-66.
126. Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M., Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25* (19), 2537-43.

127. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An overview of the Amber biomolecular simulation package. *Wires Comput Mol Sci* **2013**, *3* (2), 198-210.
128. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* **1996**, *118* (45), 11225-11236.
129. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry* **2006**, *49* (21), 6177-6196.
130. Karelson, M.; Maran, U.; Wang, Y. L.; Katritzky, A. R., QSPR and QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications. *Collect Czech Chem C* **1999**, *64* (10), 1551-1571.
131. Consonni, V.; Todeschini, R.; Wiley-Blackwell Online Books, Handbook of molecular descriptors. In *Methods and principles in medicinal chemistry 11* [Online] Wiley-VCH,: Weinheim ; New York, 2000; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC0000167881&title=Handbook%20of%20molecular%20descriptors. (visited on 11/10/2015)
132. Cherkasov, A., 'Inductive' Descriptors: 10 Successful Years in QSAR. *Current computer-aided drug design* **2005**, *1* (1), 21-42.
133. Cherkasov, A. R.; Galkin, V. I.; Zueva, E. M.; Cherkasov, R. A., New model of inductive effect: Scale of inductive electronegativity. *Zh Obshch Khim+* **1996**, *66* (5), 877-878.
134. Cherkasov, A. R.; Galkin, V. I.; Zueva, E. M.; Cherkasov, R. A., New model of inductive effect: Group inductive electronegativity. *Zh Obshch Khim+* **1996**, *66* (3), 411-415.
135. Cherkasov, A.; Jonsson, M., A new method for estimation of homolytic C-H bond dissociation enthalpies. *Journal of chemical information and computer sciences* **2000**, *40* (5), 1222-1226.
136. Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A., A new approach to the theoretical estimation of inductive constants. *J Phys Org Chem* **1998**, *11* (7), 437-447.
137. Cherkasov, A. R.; Jonsson, M.; Galkin, V., A novel approach to the analysis of substituent effects: Quantitative description of ionization energies and gas basicity of amines. *Journal of molecular graphics & modelling* **1999**, *17* (1), 28-42.
138. Cherkasov, A. R.; Galkin, V.; Cherkasov, R., "Inductive" electronegativity scale. *J Mol Struc-Theochem* **1999**, *489* (1), 43-46.
139. Cherkasov, A. R.; Galkin, V.; Cherkasov, R., "Inductive" electronegativity scale: 2. 'Inductive' analog of chemical hardness. *J Mol Struc-Theochem* **2000**, *497*, 115-121.
140. Andrews, C. G., *Scientific Vector Language as a Software Development Environment in Quantitative Structure Activity Relationships Studies*. University of New Brunswick, Faculty of Computer Science: 2003.
141. Team, R. C. R: *A language and environment for stastitical computing.* , R Foundation for Stastical Computing: Vienna, Austria, 2015.
142. Team, R. *RStudio: Integrated Development for R.*, Boston, MA, 2015.
143. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P., Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences* **2003**, *43* (6), 1947-1958.

144. Svetnik, V.; Liaw, A.; Tong, C.; Wang, T., Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. *Multiple Classifier Systems, Proceedings* **2004**, 3077, 334-343.
145. Patra, J. C.; George, N. V.; Meher, P. K., DNA Microarray Analysis Using Equalized Orthogonal Mapping. *2010 International Joint Conference on Neural Networks Ijcn* **2010**.
146. Bengio, Y.; Grandvalet, Y., No unbiased estimator of the variance of K-fold cross-validation. *J Mach Learn Res* **2004**, 5, 1089-1105.
147. Parikh, R.; Mathai, A.; Parikh, S.; Sekhar, G. C.; Thomas, R., Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* **2008**, 56 (1), 45-50.
148. Bradley, A. P., The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition* **1997**, 30 (7), 1145-1159.
149. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, 26 (16), 1668-1688.
150. Frisch, M. J., Optimizing large molecules with Gaussian 03. *Chem Listy* **2006**, 100 (4), A9-A9.
151. Labute, P., Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins-Structure Function and Bioinformatics* **2009**, 75 (1), 187-205.
152. Boyd, S., Molecular operating environment. *Chem World-Uk* **2005**, 2 (9), 66-66.
153. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry* **2003**, 24 (16), 1999-2012.
154. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of molecular graphics & modelling* **1996**, 14 (1), 33-38.
155. Lee, T. C.; Kalenius, E.; Lazar, A. I.; Assaf, K. I.; Kuhnert, N.; Grun, C. H.; Janis, J.; Scherman, O. A.; Nau, W. M., Chemistry inside molecular containers in the gas phase. *Nature chemistry* **2013**, 5 (5), 376-382.
156. Wang, J. M.; Wang, W.; Kollman, P. A., Antechamber: An accessory software package for molecular mechanical calculations. *Abstr Pap Am Chem S* **2001**, 222, U403-U403.
157. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J Phys Chem-Us* **1993**, 97 (40), 10269-10280.
158. Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling* **2006**, 25 (2), 247-260.
159. Saen-Oon, S.; Kuno, M.; Hannongbua, S., Binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-Cl TIBO complex structures: Quantum chemical calculations based on the ONIOM method. *Proteins-Structure Function and Bioinformatics* **2005**, 61 (4), 859-869.
160. Wickham, H.; Sievert, C.; SpringerLINK ebooks - Mathematics and Statistics (2016), Ggplot2 Elegant Graphics for Data Analysis. In *Use R! Ser* [Online] 2nd ed.; Springer: New York; p. 1 online resource. http://GW2JH3XR2C.search.serialssolutions.com/?sid=sersol&SS_jc=TC_024670863&title=Ggplot2. (visited on 10/01/2016)

161. Baumann, D.; Baumann, K., Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *Journal of cheminformatics* **2014**, *6* (1), 47.
162. Walton, E. B.; Vanvliet, K. J., Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2006**, *74* (6 Pt 1), 061901.
163. Fizazi, K.; Massard, C.; James, N. D.; Culine, S.; Jones, R. H.; Oksala, R.; Moilanen, A.; Aho, E.; Ravanti, L.; Kallio, P., ODM-201, a new generation androgen receptor inhibitor for castration-resistant prostate cancer: Preclinical and phase I data. *Journal of Clinical Oncology* **2013**, *31* (6).

Appendices

Appendix A QSAR modeling dataset

The QSAR modeling dataset was based upon the following mutant— anti-androgen activity class assignments where +1 is Agonist and -1 Antagonist.

S.No.	Mutant	DHT	ARN-509	Bicalutamide	Enzalutamide	Hydroxyflutamide
1	L702H	+1	-1	+1	-1	+1
2	V716M	+1	-1	+1	-1	+1
3	V731M	+1	-1	+1	-1	+1
4	W742L	+1	-1	+1	-1	+1
5	W742C	+1	-1	+1	-1	+1
6	H875Y	+1	+1	+1	+1	+1
7	H875Q	+1	-1	+1	-1	+1
8	F877L	+1	+1	-1	+1	+1
9	T878A	+1	+1	+1	+1	+1
10	T878S	+1	+1	+1	+1	+1
11	D880E	+1	-1	-1	-1	+1
12	L882I	+1	-1	+1	-1	+1
13	S889G	+1	+1	+1	-1	+1
14	D891H	+1	-1	+1	-1	+1
15	E894K	+1	-1	+1	-1	+1
16	M896V	+1	-1	+1	-1	+1
17	M896T	+1	-1	+1	-1	+1
18	E898G	+1	-1	-1	-1	+1
19	T919S	+1	-1	+1	-1	+1
20	T878A/S889G	+1	+1	+1	+1	+1
21	T878A/D891H	+1	+1	+1	+1	+1
22	F877L/T878A	+1	+1	+1	+1	+1
23	H875Y/T878A	+1	+1	+1	+1	+1
24	H875Q/T919S	+1	-1	+1	-1	+1
25	WT	+1	-1	+1	-1	+1

Appendix B List of 225 attributes screened for importance by Boruta

In Figures 3.2 and 3.3, the attributes have been hidden due to space restriction. Each of the markers for the figures on the x-axis represents 1 attribute. They have been listed below, in the direction LEFT to RIGHT on the figures.

No.	Attributes
1	r_glide_res.765_Eint
2	r_glide_res.746_dist
3	R_Nsp3.L_Osp2
4	r_glide_res.876_Eint
5	r_glide_res.750_Eint
6	r_glide_res.750_dist
7	r_glide_res.871_dist
8	r_i_docking_score
9	r_glide_res.711_dist
10	r_glide_res.704_Eint
11	r_i_glide_einternal
12	R_Nsp3.L_Nsp
13	r_glide_res.904_Eint
14	r_glide_res.875_Eint
15	r_glide_res.779_dist
16	r_glide_res.702_dist
17	r_glide_res.743_Eint
18	R_Nsp2.L_Osp2
19	r_glide_res.784_dist

20	R_Csp3.L_Nsp2
21	R_Osp3.L_Nsp2
22	R_Nsp2.L_Nsp2
23	R_Ssp3.L_Nsp2
24	r_glide_res.765_dist
25	r_i_glide_rmsd_to_input
26	r_glide_res.703_dist
27	r_glide_res.701_dist
28	r_glide_res.781_Eint
29	R_Csp2.L_Nsp2
30	R_Osp2.L_Nsp2
31	R_Nsp3.L_Nsp2
32	R_Nsp3.L_Csp
33	r_glide_res.874_dist
34	X878.Sigma_AA_L
35	r_glide_res.708_Eint
36	r_glide_res.788_Eint
37	R_Nsp2.L_Fsp3
38	R_Ssp3.L_Fsp3
39	r_glide_res.788_dist
40	r_glide_res.711_Eint
41	r_glide_res.705_dist
42	r_glide_res.896_Eint
43	X891.Rs_AA_L
44	r_glide_res.771_dist
45	r_glide_res.742_dist

46	R_Csp2.L_Csp
47	r_glide_res.877_dist
48	R_Osp2.L_Fsp3
49	r_glide_res.707_Eint
50	r_glide_res.881_Eint
51	X731.Sigma_AA_L
52	R_Ssp3.L_Osp2
53	R_Ssp3.L_Ssp2
54	r_glide_res.702_Eint
55	R_Csp2.L_Fsp3
56	r_glide_res.701_Eint
57	r_glide_res.708_dist
58	r_i_glide_evdw
59	R_Ssp3.L_Nsp
60	X891.Abs_Rs_AA_L
61	r_glide_res.896_dist
62	R_Nsp2.L_Ssp2
63	X877.878.Sigma_AA_L
64	r_glide_res.877_Eint
65	r_glide_res.769_dist
66	r_glide_res.706_dist
67	X702.Sigma_AA_L
68	R_Nsp3.L_Fsp3
69	r_i_glide_energy
70	r_glide_res.712_dist
71	r_glide_res.707_dist

72	X878.891.Sigma_AA_L
73	X896.Sigma_AA_L
74	r_glide_res.881_dist
75	R_Csp3.L_Ssp2
76	R_Csp2.L_Osp2
77	X878.889.Sigma_AA_L
78	r_glide_res.742_Eint
79	X877.878.Rs_AA_L
80	R_Osp2.L_Ssp2
81	X878.Rs_AA_L
82	r_glide_res.710_Eint
83	X742.Sigma_AA_L
84	r_glide_res.876_dist
85	X889.Sigma_AA_L
86	X878.889.Rs_AA_L
87	R_Osp3.L_Nsp
88	r_glide_res.712_Eint
89	r_glide_res.892_dist
90	R_Csp2.L_Ssp2
91	r_glide_res.870_dist
92	Sigma_L_R
93	R_Csp2.L_Nsp
94	R_Osp2.L_Osp2
95	X880.Sigma_AA_L
96	X702.Abs_Rs_AA_L
97	r_glide_res.899_dist

98	X716.Sigma_AA_L
99	r_glide_res.705_Eint
100	r_glide_res.747_Eint
101	R_Csp3.L_Osp2
102	r_glide_res.703_Eint
103	X877.Sigma_AA_L
104	X880.Abs_Rs_AA_L
105	r_glide_res.879_dist
106	X894.Sigma_AA_L
107	r_glide_res.713_dist
108	R_Osp2.L_Csp
109	X889.Rs_AA_L
110	R_Csp3.L_Osp3
111	X877.Rs_AA_L
112	r_glide_res.874_Eint
113	r_glide_res.704_dist
114	r_glide_res.893_Eint
115	r_glide_res.873_Eint
116	r_i_glide_ligand_efficiency_ln
117	R_Csp3.L_Fsp3
118	R_Nsp2.L_Osp3
119	R_Ssp3.L_Osp3
120	r_glide_res.784_Eint
121	r_i_glide_ecoul
122	R_Osp2.L_Nsp
123	r_glide_res.769_Eint

124	r_glide_res.771_Eint
125	R_Nsp2.L_Csp
126	X898.Sigma_AA_L
127	X875.Rs_AA_L
128	R_Ssp3.L_Csp3
129	r_glide_res.747_dist
130	r_glide_res.709_dist
131	R_Osp2.L_Osp3
132	X891.Sigma_AA_L
133	R_Nsp3.L_Csp3
134	r_glide_res.875_dist
135	X880.Rs_AA_L
136	R_Osp3.L_Ssp2
137	R_Nsp2.L_Ssp3
138	X878.891.Rs_AA_L
139	R_Nsp3.L_Csp2
140	R_Osp3.L_Ssp3
141	R_Nsp2.L_Csp3
142	R_Ssp3.L_Csp
143	r_glide_res.710_dist
144	R_Nsp3.L_Osp3
145	R_Csp2.L_Osp3
146	R_Nsp3.L_Nsp3
147	r_glide_res.904_dist
148	R_Ssp3.L_Ssp3
149	X875.Abs_Rs_AA_L

150	X878.Abs_Rs_AA_L
151	R_Osp2.L_Ssp3
152	Q_L_R
153	R_Csp2.L_Ssp3
154	R_Osp3.L_Csp
155	r_glide_res.706_Eint
156	R_Csp3.L_Ssp3
157	R_Osp3.L_Osp3
158	r_glide_res.900_Eint
159	r_glide_res.878_Eint
160	r_i_glide_ligand_efficiency
161	r_glide_res.900_dist
162	r_glide_res.781_dist
163	r_glide_res.744_Eint
164	X889.Abs_Rs_AA_L
165	X878.891.Abs_Rs_AA_L
166	R_Nsp3.L_Ssp3
167	X702.Rs_AA_L
168	R_Osp3.L_Osp2
169	r_glide_res.744_dist
170	R_Csp3.L_Csp3
171	R_Osp2.L_Csp3
172	R_Csp2.L_Csp3
173	Abs_Rs_L_R
174	Rs_L_R
175	r_glide_res.892_Eint

176	r_glide_res.899_Eint
177	X877.878.Abs_Rs_AA_L
178	r_glide_res.893_dist
179	X878.889.Abs_Rs_AA_L
180	r_glide_res.713_Eint
181	R_Nsp2.L_Csp2
182	R_Ssp3.L_Csp2
183	r_glide_res.743_dist
184	R_Nsp2.L_Nsp
185	X898.Rs_AA_L
186	X875.Sigma_AA_L
187	r_glide_res.709_Eint
188	R_Csp3.L_Csp
189	R_Csp3.L_Csp2
190	X894.Rs_AA_L
191	r_glide_res.879_Eint
192	X877.Abs_Rs_AA_L
193	R_Csp3.L_Nsp
194	X882.Abs_Rs_AA_L
195	r_glide_res.873_dist
196	r_glide_res.746_Eint
197	X882.Rs_AA_L
198	R_Nsp2.L_Nsp3
199	X882.Sigma_AA_L
200	R_Csp2.L_Csp2
201	X898.Abs_Rs_AA_L

202	R_Osp3.L_Csp3
203	R_Ssp3.L_Nsp3
204	r_glide_res.878_dist
205	r_glide_res.779_Eint
206	R_Osp2.L_Csp2
207	X894.Abs_Rs_AA_L
208	R_Csp3.L_Nsp3
209	r_glide_res.870_Eint
210	R_Nsp3.L_Ssp2
211	X716.Abs_Rs_AA_L
212	X742.Abs_Rs_AA_L
213	r_glide_res.871_Eint
214	r_i_glide_emodel
215	R_Csp2.L_Nsp3
216	R_Osp3.L_Fsp3
217	X731.Abs_Rs_AA_L
218	X742.Rs_AA_L
219	R_Osp2.L_Nsp3
220	X731.Rs_AA_L
221	R_Osp3.L_Csp2
222	R_Osp3.L_Nsp3
223	X896.Rs_AA_L
224	X716.Rs_AA_L
225	X896.Abs_Rs_AA_L