

**MOLECULAR MODELING STUDIES ON HIV-1
INHIBITORS AND THEIR POTENTIAL USE AS
ANTICANCER AGENTS**

OLAYIDE ADEBIMPE ARODOLA

213569288



A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal, Westville, in
fulfilment of the requirements of the degree of Master of Medical Sciences

Supervisor

Prof. Mahmoud Soliman

KwaZulu-Natal

2014

**MOLECULAR MODELING STUDIES ON HIV-1
INHIBITORS AND THEIR POTENTIAL USE AS
ANTICANCER AGENTS**

2014

ARODOLA OLAYIDE ADEBIMPE

213569288

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville Campus, for the degree of Master of Medical Sciences (Pharmaceutical Chemistry).

This is the thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary.

This is to certify that the contents of this thesis are the original research work of Miss Olayide Adebimpe Arodola.

As the candidate's supervisor, I have approved this thesis for submission.

Supervisor:

Signed: ----- Name: **Prof. Mahmoud E. Soliman** Date: -----

ABSTRACT

Acquired Immunodeficiency Syndrome (AIDS), currently regarded as one of the deadliest diseases, is a disease of the human immune system caused by the Human Immunodeficiency Virus (HIV). This dissertation addresses two classes of HIV-1 inhibitors: (i) integrase and (ii) protease inhibitors. With the first class, a 2D-QSAR study was carried out on compounds from a variety of structural classes; 40 diketo acid and carboxamide derivatives; possessing integrase inhibitory activity. This study investigated the relationship between molecular properties and HIV-1 integrase inhibitor activities and established accurate QSAR predictive model using the Genetic Function Algorithm (GFA) statistical model. The logarithmic inverse values of IC_{50} (μM) and physicochemical parameters represent the dependent variable and independent variable, respectively. Results demonstrated that the radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with the correlation coefficient value of 0.820 and play an important role in HIV-1 integrase inhibition.

With the second class, the binding affinities of some FDA-approved HIV-1 protease inhibitors, which were reported to possess anticancer activities, were estimated. The findings proposed here may alter perceptions about how NFV binds to the human Hsp90; the protein responsible for the overexpression of HER2+ breast cancer; since it has only been reported to inhibit NSCLC and a collection of yeast strains. A human Hsp90 homologue was built due to the lack of a full X-ray crystal structure of the human Hsp90 on protein data bank. The Ramachandran plot showed the validity of the human Hsp90 homologue where 98% of all residues, including the active site residues, were in the favoured region and 99.8% were in the allowed region. The NTD active acid residues were found to be Leu43, Asn46, Lys53, Ile91, Asp97, Met93, Asn101, Ser108, Gly109, Phe133 and Thr179. The obtained active site residues for the human Hsp90 homologue CTD were Gln523, Val534, Ser535, Lys538, Thr595, Tyr596, Gly597, Trp598 and Met602.

The system stability and overall convergence of simulations were evaluated. The RMSD of all nine PIs did not exceed 2Å and the system stabilised after 1000 ps and 1800 ps MD

simulation at the NTD and CTD, respectively. The fluctuations of potential energies at the NTD were <2000 kcal/mol for 5 ns of MD simulation and CTD show that the fluctuations of the potential energy to be ≤ 8000 kcal/mol. The free binding energy of NFV was -83.03 kcal/mol at the NTD and -39.3 kcal/mol at the CTD. This value shows a significant difference (~ 43.73 kcal/mol) between the interaction energy at the NTD and CTD. Energy decomposition analysis at the NTD and CTD show that these two active sites have major energy contributions from their respective active site residues.

This study is of great importance to medicine as it predicts the biological activity of some potent HIV-1 IN and investigates the potential use of the current HIV-1 PR drugs as anticancer agents.

DECLARATION 1 – PLAGIARISM

I, **Olayide Adebimpe Arodola**, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original work.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a. Their words have been re-written, but the general information attributed to them has been referenced.
 - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references section.

A detail contribution to publications that form part and/or include research presented in this thesis is stated (include publications submitted, accepted, in *press* and published).

Signed

DECLARATION 2 - LIST OF PUBLICATIONS

1. Arodola, O. A., Radha Charan, D., and Mahmoud, E. S. S. (2014) QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor, *Lett Drug Des Discov* 11, 618-627.

Contribution:

Olayide Adebimpe Arodola: Author- contributed to the project by performing all literature reviews, experimental work, and data analysis, interpretation of the results as well as manuscript preparation and writing.

Radha Charan Dash: Provided assistance with technical support

Mahmoud E. Soliman: Supervisor

Appendix A.3: PDF version of the publication

2. Could the FDA-approved anti-HIV drugs be promising anticancer agents? Answer from extensive molecular dynamics analyses (**Submitted to Chemical Biology & Drug Design**)

Contribution:

Olayide Adebimpe Arodola: Contributed to the project by performing 50% literature review, experimental work, and data analysis, interpretation of results, manuscript preparation and writing.

Sbongile Mbatha: Contributed to the project by performing 50% experimental work, data analysis, manuscript preparation and writing.

Mahmoud E. Soliman: Supervisor

RESEARCH OUTPUT

A. PUBLICATIONS

1. Arodola, O. A., Radhan Charan, D., and Mahmoud, E. S. S. (2014) QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor, *Lett Drug Des Discov* 11, 618-627.
2. Could the FDA-approved anti-HIV drugs be promising anticancer agents? Answer from extensive molecular dynamics analyses (Manuscript submitted and under review)

B. CONFERENCES

1. Oral presentation “QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor” – CHPC national conference, Cape Town, 1-5 December, 2013.
2. Poster presentation “QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor” – 3’s company pharmacy conference, Cape Town, 3-6 October, 2013.

DEDICATION

This work is dedicated to God the giver of life, thanking him for his unfailing grace, faithfulness and unequalled love towards me.

This is also dedicated to my parents and the Arodola family for their love and support throughout my studies and every other day of my life as well as for teaching me many life skills that has contributed to my success. They deserve more than one dedication for being such a supportive family.

And lastly, this work is dedicated to Oladoyinbo Segun, an amazing man, without him, there will be no thesis. His support, perseverance, confidence, strength and optimism are catching. “Forever and always *plus a bit more and an extra day*”.

ACKNOWLEDGEMENTS

- I wish to express my sincere gratitude and a gigantic thanks to my supervisor Dr. Mahmoud E. Soliman for being so supportive, listening, giving the best advice, guidance, patience and moral support during the course of my degree. He taught me so much pharmaceutical chemistry and countless scientific techniques, his knowledge is awe inspiring and his teaching methodologies are beyond comparison.
- My thanks also go out to the postdocs, Dr. Sarentha Chetty, Dr. Ramesh and Dr. Radha Charan Dash. It is a privilege working with you.
- A big thank you to the UKZN molecular modeling and drug design research group (2013/2014 group) for sharing their research knowledge with me. You all are highly appreciated. It's been wonderful working with you all.
- To the friends that have positively contributed to my progress at one stage or another, a big hug goes to you all.
- My profound gratitude goes to CHPC for their resources and technical support
- My appreciation goes to UKZN college of Health Sciences for financial support throughout the course of my study.

TABLE OF FIGURES

Figure 2.1. A model structure of the HIV-1 virus.....	13
Figure 2.2. The HIV-1 replication cycle.....	14
Figure 2.3. HIV-1 Integrase inhibitors.....	18
Figure 2.4. FDA-approved HIV-1 protease inhibitors.....	19
Figure 2.5. The structure of the HIV-1 integrase enzyme.....	20
Figure 2.6. Structures of the three domains of HIV-1 integrase.....	21
Figure 2.7. DNA cutting and joining steps in HIV integration.....	22
Figure 2.8. Structure of HIV-1 protease in complex with an inhibitor.....	23
Figure 2.9. The model structure of a normal and cancer-infected breast cell.....	24
Figure 2.10. The crystal structure of Hsp90.....	25
Figure 3.1. A model of a two-dimensional potential energy surface.....	37
Figure 3.2. Schematic representation of the protocol followed during construction of the human Hsp90 homologue.....	44
Figure 4.1. Predicted pIC ₅₀ against experimental pIC ₅₀ for equation 1.....	61
Figure 4.2. Predicted pIC ₅₀ against experimental pIC ₅₀ for equation 2.....	61
Figure 4.3. Predicted pIC ₅₀ against experimental pIC ₅₀ for equation 3.....	62
Figure 4.4. Histogram of residual values obtained for equation 1.....	64
Figure 4.5. Histogram of residual values obtained for equation 2.....	65
Figure 4.6. Histogram plot of residual values obtained for equation 3.....	65
Figure 4.7. The graph of the variable usage against generation number.....	66
Figure 5.1. The crystal structure of Hsp90 alpha and beta chain.....	73
Figure 5.2. 2D structures of the nine FDA-approved HIV-1 protease inhibitors.....	76
Figure 5.3. Nelfinavir-human Hsp90 homologue interaction at the NTD and CTD.....	80
Figure 5.4. Generated homology model of the human Hsp90β.....	84
Figure 5.5. The human Hsp90 homologue docked with NFV at NTD and CTD.....	85
Figure 5.6. Comparative RMSD and potential energy plot of all ligands binding at the NTD and CTD.....	86
Figure 5.7. Comparative RMSF plot of all ligands binding at the NTD and CTD.....	88
Figure 5.8. Residues that contributed to the human Hsp90-nelfinavir binding at the NTD and CTD.....	92

Figure 6. Comparative RMSF plot for PIs at the human Hsp90 homologue NTD and CTD.....	104
Figure 7a. Comparative RMSD plot for PIs at the human Hsp90 homologue NTD and CTD.....	107
Figure 7b. Comparative potential energy plot for PIs at the human Hsp90 homologue NTD and CTD.....	109
Supplementary Figure-S 1. The multiple sequence alignment result	112
Supplementary Figure-S 2. Ramachandran plot for the human Hsp90 homologue	114
Supplementary Figure-S 3. Ligand-enzyme interaction at the human Hsp90 homologue NTD and CTD.....	116
Supplementary Figure-S 4. Comparative per-residue decomposition energy at the human Hsp90 NTD and CTD	126

TABLE OF TABLES

Table 4.1. Structures and biological activity of training and test set	54
Table 4.2. List of descriptors used in this study.....	57
Table 4.3. The 3 different equations derived from the QSAR model	58
Table 4.4. Experimental pIC ₅₀ and GFA predicted pIC ₅₀ for training set.....	62
Table 4.5. Experimental pIC ₅₀ and GFA predicted pIC ₅₀ for test set.....	63
Table 5.1. Binding free energies of the FDA-approved protease inhibitors against the human Hsp90 homologue.	90

LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
Akt/Pkb	Protein Kinase B
APV	Amprenavir
ART	Antiretroviral Therapy
ATP	Adenosine Triphosphate
ATV	Atazanavir
β	Beta
CA	Alpha Carbon
CADD	Computer-Aided Drug Design
CAMD	Computer-Assisted Molecular Design
CCR5	C-C Chemokine Receptor 5
CD4	Cluster of Differentiation 4
CDC	Centre for Disease Control
C-SA	Subtype C-South African
CTD	C-Terminal Domain
CXCR4	C-X-C Chemokine Receptor type 4
DKA	Diketo Aryl
DNA	Deoxyribonucleic Acid
DRV	Darunavir
ED ₅₀	50% Effective Dose
ER	Estrogen Receptor
FDA	Food and Drug Agency (USA)
FDT	Fluctuation-Dissipation Theorem
FEP	Free Energy Perturbation
GFA	Genetic Function Algorithm
GLOBOCAN	International Agency for Research on Cancer
HAART	Highly Active Antiretroviral Therapy
HER2+	Human Epidermal Growth Factor 2- positive

LIST OF ABBREVIATIONS

HIV-1	Human Immunodeficiency Syndrome type 1
HSP	Heat Shock Protein
HUMO	Highest Un-occupied Molecular Orbital
IC ₅₀	50% Inhibitory Concentration
IDV	Indinavir
IN	Integrase
INSTI	Integrase Strand Transfer Inhibitor
LOF	Lack of Fit
LOO	Leave-one-out method
LPV	Lopinavir
LR	Linear Response
LUMO	Lowest Un-occupied Molecular Orbital
MD	Molecular Dynamics
MLR	Multiple Linear Regression
MM	Molecular Mechanics
MM/GB-SA	Molecular Mechanics General Born-Surface Area
MM/PB-SA	Molecular Mechanics Poisson Boltzman-Surface Area
mRNA	Messenger Ribonucleic Acid
NFV	Nelfinavir
NMR	Nuclear Magnetic Resonance
NNRTI	Non-nucleoside Reverse Transcriptase Inhibitor
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NSCLC	Non-small cell lung carcinoma
NTD	N-Terminal Domain
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDB	Protein Data Bank
PES	Potential Energy Surface
PIs	Protease Inhibitors
PLS	Partial Least Square
PR	Protease

LIST OF ABBREVIATIONS

Q ²	Cross-validated Correlation Coefficient
QSAR	Quantitative Structural Activity Relationship
QSPR	Quantitative Structure Property Relationship
R ²	Correlation Coefficient
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
RTV	Ritonavir
SQV	Saquinavir
TI	Thermodynamic Integration
TPV	Tiprenavir
UNAIDS	The Joint United Nations Program on HIV/AIDS

TABLE OF CONTENTS

MOLECULAR MODELING STUDIES ON HIV-1 INHIBITORS AND THEIR POTENTIAL USE AS ANTICANCER AGENTS.....	1
MOLECULAR MODELING STUDIES ON HIV-1 INHIBITORS AND THEIR POTENTIAL USE AS ANTICANCER AGENTS.....	i
ABSTRACT.....	i
DECLARATION 1 – PLAGIARISM.....	iii
DECLARATION 2 - LIST OF PUBLICATIONS	iv
RESEARCH OUTPUT.....	v
DEDICATION.....	vi
ACKNOWLEDGEMENTS	vii
TABLE OF FIGURES.....	viii
LIST OF ABBREVIATIONS	ix
TABLE OF CONTENTS	xi
CHAPTER 1	1
1.1. Background to and rationale for this study	1
1.2. Aims and objectives of this study	3
1.3. Novelty and significance of this study	4
1.4. Overview of this thesis.....	6
References:.....	8
CHAPTER 2	12
2.1. Introduction	12
2.2. A brief history of HIV/AIDS	12
2.3. The HIV-1 virus	13
2.4. The HIV-1 life cycle	14

2.5.	Current antiretroviral treatments for HIV-1 virus infection.....	16
2.5.1.	Reverse transcriptase inhibitors	17
2.5.2.	Fusion inhibitors	17
2.5.3.	Integrase inhibitors.....	17
2.5.4.	Protease inhibitors.....	18
2.6.	HIV-1 viral enzymes	20
2.6.1.	HIV-1 integrase enzyme	20
2.6.2.	HIV-1 protease enzyme	22
2.7.	Brief Introduction to Drug Repositioning	23
2.8.	The use of HIV-1 protease inhibitors as potential anticancer agents.....	24
	References:.....	27
CHAPTER 3		35
3.1.	Introduction to Computational Chemistry	35
3.2.	Schrödinger's equation.....	35
3.3.	Born-Oppenheimer approximation	36
3.4.	Potential Energy Surface	37
3.5.	Molecular Mechanics	37
3.5.1.	Force fields.....	38
3.6.	Molecular Dynamics simulations.....	39
3.7.	Approaches for estimating binding affinities.....	40
3.7.1.	Molecular docking	40
3.7.2.	Binding free energy calculations	42
3.7.3.	Entropy calculations.....	42
3.8.	Molecular modeling tools used in this study.....	43
3.8.1.	Homology modeling	43

3.8.2. Per-residue energy decomposition analysis	44
3.8.3. Quantitative structural activity relationship (QSAR)	45
References:	47
CHAPTER 4	51
Abstract:	51
Introduction	52
Methods	56
Calculation of pIC50	56
Calculation of molecular descriptors	57
Conclusion	57
Acknowledgement	67
Conflict of Interest	67
References:	68
CHAPTER 5	71
Abstract	71
1. Introduction	72
2. Computational Methodology	77
2.1. Homology modeling of human Hsp90 protein structure	77
2.2. Defining the active site residues in the Hsp90 homology model	77
2.2.1. N-terminal Domain (NTD)	78
2.2.2. C-terminal domain (CTD)	78
2.3. Building Hsp90-HIV protease inhibitor complexes	81
2.4. Molecular Dynamics Simulations	81
2.5. Thermodynamic Calculation	82
3. Results and Discussion	83

3.1. The Human Hsp90 homology model	83
3.2. Molecular Dynamics (MD) simulations.....	85
3.2.1. Post-dynamic analyses:.....	85
3.2.2. MM/GBSA free binding energy calculations	89
3.2.3. Per-residue interaction energy decomposition analysis	91
4. Conclusion	93
Supplementary Materials	93
Acknowledgement	93
Conflicts of Interest.....	93
Chapter 6	99
6.1. General Conclusion.....	99
6.2. Recommendations and Future Studies	101
References:.....	103
APPENDICES	104
Appendix 1. Supplementary material for Chapter 5	104
Appendix 2. Input files for docking the human Hsp90 homologue complexes at NTD	129
Appendix 3. Input files for docking the human Hsp90 homologue complexes at CTD	129
Appendix 4. PDF version of the published paper	130

CHAPTER 1

1.1. Background to and rationale for this study

This chapter discusses the background and novelty of research projects related to HIV-1 infection and HER2+ breast cancer. In Africa, the AIDS epidemic has achieved high figures. Of the 42 million individuals infected with HIV worldwide, 30 million are in Africa (1) and currently, approximately 5.7 million are in South Africa (2). An important reason that might contribute to this high figure is that HIV-1 subtypes prevalent in sub-Saharan Africa have not attracted sufficient attention from researchers compared to the subtypes in North America and Western Europe (3). Out of the ten subtypes of HIV-1, subtype C has a 95% occurrence rate of HIV infections in South Africa (4). The social and economic pressure exerted by HIV pandemic in South Africa has evoked an international response towards aiding HIV research in prevention and treatments (5).

Acquired immunodeficiency syndrome (AIDS) is a disease caused by the Human Immunodeficiency Virus (HIV). The ability of the HIV-1 to evade the immune system coupled with its replication and increase drug resistance makes the treatment of the disease difficult. There are two strains of HIV; HIV-1 and HIV-2 (6, 7). Various enzymes play key role in the replication of HIV-1 virus such as; integrase, protease, reverse transcriptase and thus they are the main targets for the current HIV drugs (8). The HIV-1 virus progresses by first invading the host DNA. In the first step of the integration process, two nucleotides are removed from the 3'-end of the viral DNA by a reaction termed 3'-end processing. Cleavage occurs at the 3'-side of the CA dinucleotide that is conserved among retroviruses, retro-transposons, and several DNA transposons, both in prokaryotes and eukaryotes (9). The integration of HIV-1 viral DNA into the host DNA, involves a series of reactions which are vital in the replication cycle of HIV-1 (10) and other retroviruses (11, 12).

Available data reveal that there are numerous compounds undergoing screening, which are yet to be approved as HIV-1 integrase inhibitor by the Food and Drug Agency (FDA)

(13-15). An important fact is that screening of a large library of ligands requires a short period of time in evaluating a single compound. This obvious problem brought forth the use of computers in drug design to help speed up the screening of lead compounds by using the quantitative structural activity relationship (QSAR) approach. This method attempts to correlate the molecular properties to the physicochemical properties of a set of structures to predict the biological activities of other compounds, thereby reducing the time spent on synthesizing a large library of compounds. This study investigated compounds observed to inhibit HIV-1 integrase enzymes to obtain more insight into their inhibitory profile. The strand transfer data of 40 diketo acid and carboxamide derivatives were retrieved from the literature (16-18) and analyzed using the QSAR approach (19).

Furthermore, the HIV-1 protease enzyme, which slices the *gag* and *pol* nonfunctional polypeptide into functional proteins essential for the development of mature infectious HIV particles (20), is considered a crucial target for designing of HIV-1 inhibitors (21). Protease inhibitors are one of the numerous success stories of medicinal chemistry and recently some protease inhibitors have been reported as potential anticancer agents through the process called ‘repositioning’ (22-24).

Cancer is an abnormal proliferation of cells that can lead to death and these cells have the potential to metastasize (25). Breast cancer is a malignant tumor, which starts in the cells of the breast and it is commonly diagnosed among women, although not limited to women. Invasive breast cancers are typically initially categorized into three subtypes based on the cellular expression of certain receptors. These are the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2). Survival rates are based on the phenotype expression for each patient, though, ER+ and PR+ patients tend to have a better prognosis compared to triple negative patients (ER- PR- HER2-) who have the poorest prognosis (26).

About twenty-eight years ago, HER2, one of a family of four membrane tyrosine kinases was reported to be “amplified” in 20-25% of human breast cancer cells (27). Two years later, this amplification was found to be important in the pathogenesis and the

advancement of human breast cancer (28). Since that time, the detection of HER2 has become a routine prognostic and predictive factor in breast cancer. The HER2+ breast cancer has been reported to overexpress heat shock proteins (HSPs) in the proliferating cell (29). Cancer cells have been reported to be dependent on a particular protein (30), Hsp90, one of a group of molecular chaperones responsible for the function of other client proteins (31); thus, this represents a molecular target in the treatment of HER+ breast cancer.

For the work described herein, we anticipate and investigate the anticancer property of the recent FDA-approved HIV-1 protease inhibitors (24) by performing eighteen 5 ns molecular dynamics simulation and post-dynamic analysis, such as RMSD, RMSF, free binding energy calculations and per-residue energy decomposition analysis to help us gain more insight into their binding mode and potential use as HER2+ breast cancer inhibitors.

1.2. Aims and objectives of this study

The major aims of this study were to investigate the drugs that are known to inhibit HIV-1 integrase and HER2+ breast cancer. These aims are outlined below:

1. To investigate the inhibitory profile of a number of diketo acid and carboxamide derivatives as HIV-1 integrase inhibitor using the QSAR approach. To accomplish this, the following specific objectives were outlined:
 - 1.1. To identify and acquire sets of structures and their experimental biological activities (IC_{50}) from literature.
 - 1.2. To divide the sets of structures into training and test sets and predict the biological activities of the test set.
 - 1.3. To develop QSAR equations for the predicted biological activities using the GFA statistical model.
 - 1.4. To compare the experimental biological activities against the QSAR predicted biological activities and indicate the physicochemical properties that contribute to the highest correlation coefficient R^2 .

2. To provide a molecular understanding on the effectiveness of the clinically reported FDA-approved PIs against HER2+ breast cancer. To accomplish this, the following specific objectives were outlined:
 - 2.1. To build a homology model due to the absence of the human Hsp90 complete crystal structure in the protein data bank.
 - 2.2. To validate the human homologue by plotting a Ramachandran plot to determine the outliers.
 - 2.3. To carry out docking on the human Hsp90 homologue NTD and CTD to nine protease inhibitors and use the docked 18 complexes for subsequent simulation protocols.
 - 2.4. To perform MD simulations, post-dynamic calculations and binding free energy calculations on the homologue-drug complexes.
 - 2.5. To estimate the contribution of each amino acids towards the overall binding to each drug by calculating the ligand-enzyme interaction and per-residue energy decomposition analysis.
 - 2.6. To perform a comparative study on which of the Hsp90 terminals binds better to each drug since it was reported that the CTD also have bioactive residues.

1.3. Novelty and significance of this study

Almost every QSAR study consists of several QSAR models, with several QSAR studies, including 2D-, 3D- and even 4D-QSAR analyses, have been reported over the past 15 years (32-38) on a variety of integrase inhibitors (IN) to clarify the quantitative correlations between the chemical structures of IN and their biological activities (39). The GFA method, among other statistical methods, is used for identifying optimal solutions to a problem where the possible solution space is too large to be exhaustively enumerated (40-42). In spite of the GFA method been reported in some published data to be an effective tool for performing both QSAR and QSPR (43-45), the QSAR study on HIV-1 integrase inhibitors, such as carboxamide and diketo acids using this GFA method has few reviews. The predictive ability of a QSAR model is usually measured by a cross-

validated Q^2 value and a correlation coefficient R^2 . Using these values as ranking criteria, result show the best QSAR models (see chapter 4) from QSAR studies of IN inhibitors with experimental biological activities (16-18).

Furthermore, it was reported that Hsp90 contains a highly conserved ATP binding domain near its N-terminus (46). This Hsp90 ATP binding site has been under intense pharmaceutical investigation, as the majority of known Hsp90 inhibitors binds to this site in competition with ATP (47). It has also been found that the C-terminus show some promising activities at its binding pocket (47), but this theory is yet to be experimentally proven or thoroughly validated through computational methods.

The study done by Gills *et al.* 2007 found that nelfinavir, which is one of the clinically approved HIV protease inhibitors, inhibited growth of a wide variety of cancer cell types at concentrations that have been achieved in patients infected with HIV through the inhibition of Akt signaling pathway (23). Studies also hypothesize that nelfinavir binds Hsp90 at the C-terminal domain and induces conformational changes in the protein; this is a different mechanism compared to other Hsp90 inhibitors (24). However, both of these studies did not show the precise mode of interaction between nelfinavir and Hsp90 and there is no understanding on how HIV-protease inhibitors play roles in HER2+ breast cancer inhibition. Both studies reported NSCLC and numerous yeast strains against HIV-1 PIs, however, there is no information regarding the effect of HIV-1 PIs on human HER2+ breast cancer. This gap of information forms the basis of this study wherein a comprehensive study of the inhibitory activity of the current FDA approved HIV-1 protease inhibitors against HER2+ breast cancer is presented. In this work, the binding site of the human Hsp90 homologue to these protease inhibitors is identified and validated. This identification and validation encompasses a robust, integrated *in-silico* approach using computational tools like molecular dynamic simulations, binding free energy calculations and post-dynamic analyses.

1.4. Overview of this thesis

This thesis is divided into six chapters, this one included:

Chapter 2: It provides a general overview on the HIV/AIDS epidemic and therapy. The chapter starts with a historical background on HIV/AIDS epidemic then some updated statistics on the number of HIV-infected people worldwide and in Africa. The chapter also highlights many aspects such as HIV virus structure; life cycle and the essential enzymes required for virus maturation and HIV drug targets. The HIV integrase and protease enzymes, as crucial drug targets and the main focus of this work, are then addressed in details, including its structure, inhibitor design strategies and the current FDA-approved protease inhibitors. A brief overview on drug repositioning and the potential use of protease inhibitors as potential HER2+ breast cancer inhibitor ends the chapter.

Chapter 3: This chapter details a general introduction to computational chemistry and different molecular modeling and simulation techniques as well as their common applications. Some theoretical descriptions of the computational methods have been included, followed by various computational tools used in HIV research with main focus on molecular dynamics simulations, quantum mechanics, molecular mechanics, molecular docking, binding free energy calculations and QSAR. Some significant previously reported computational studies on HIV-1 IN and potential use of PIs as HER2+ breast cancer inhibitors were also included at the end of the chapter.

Chapter 4: (Published work – this chapter is presented in the required format of the journal and is the final revised accepted version)

This is a research paper titled “QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor” and it was published in the Journal of Letters in drug design and discovery (48). It addresses the objectives 1.1, 1.2, 1.3 and 1.4.

Chapter 5: (Submitted work – this chapter is presented in the required format of the journal and is the final revised submitted version)

This is a research paper from this study. The paper titled “Could the FDA-approved anti-HIV drugs be promising anticancer agents?” It addresses the objectives 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6.

Chapter 6: This expounds the concluding remarks of the entire thesis and future work plans.

References:

1. UNAIDS.
(http://data.unaids.org/pub/GlobalReport/2008/jc1510_2008_global_report_pp211_234_en.pdf) 20 Jan 2014.
2. UNAIDS.
(http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/201309_epi_core_en.pdf) 20 Jan 2014.
3. AIDS global fact sheet. (2013)
http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_unaids_global_report_2012_with_annexes_en.pdf 03 Jun. 2014.
4. Mosebi, S., Morris, L., Dirr, H. W., and Sayed, Y. (2008) Active-site mutations in the south african human immunodeficiency virus type 1 subtype c protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study, *J. Virol.* 82, 11476-11479.
5. Scott, D. (2002) The impact of HIV/AIDS on rural households and land issues in Southern and Eastern Africa,
<ftp://ftp.fao.org/docrep/nonfao/ad696e/ad696e600.pdf> June 2014.
6. Clavel, F., Guetard, D., Brunvezinet, F., Chamaret, S., Rey, M. A., Santosferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L., and Montagnier, L. (1986) Isolation of a new human retrovirus from west-african patients with AIDS, *Sci* 233, 343-346.
7. Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L., and Alizon, M. (1986) Molecular-cloning and polymorphism of the human immune-deficiency virus type-2, *Nature* 324, 691-695.
8. Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Curr. Drug Targets: Infect. Disord.* 2, 17-32.
9. Craigie, R. (2001) HIV integrase, a brief overview from chemistry to therapeutics, *J. Biol. Chem.* 276, 23213-23216.
10. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) Disruption of the terminal base pairs of retroviral DNA during integration, *Genes Dev* 11, 371-382.
11. Asante-Appiah, E., and Skalka, A. M. (1997) Molecular mechanisms in retrovirus DNA integration, *Antiviral Res.* 36, 139-156.
12. Hindmarsh, P., and Leis, J. (1999) Retroviral DNA integration, *Microbiol. Mol. Biol. Rev.* 63, 836-843.
13. David, C. A., Middleton, T., Montgomery, D., Ben, L. H., Kati, W., Molla, A., Xuei, X. L., Warrior, U., Kofron, J. L., and Burns, D. J. (2002) Microarray compound screening (mu ARCS) to identify inhibitors of HIV integrase, *J. Biomol. Screen.* 7, 259-266.
14. Ren-Rong, T., Qing-Jiao, L., and Xti-Fin, C. (2007) Current status of targets and assays for Anti-HIV drug screening, *Virol. Sin.* 22, 476-485.
15. Barreca, M. L., De Luca, L., Ferro, S., Rao, A., Monforte, A. M., and Chimirri, A. (2006) Computational and synthetic approaches for the discovery of HIV-1 integrase inhibitors, *Arkivoc*, 224-244.

16. Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Paz, O. G., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., and Rowley, M. (2008) Discovery of Raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, *J. Med. Chem.* *51*, 5843-5855.
17. Wai, J. S., Egbertson, M. S., Payne, L. S., Fisher, T. E., Embrey, M. W., Tran, L. O., Melamed, J. Y., Langford, H. M., Guare, J. P., Zhuang, L. G., Grey, V. E., Vacca, J. P., Holloway, M. K., Naylor-Olsen, A. M., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Schleif, W. A., Gabryelski, L. J., and Young, S. D. (2000) 4-aryl-2,4-dioxobutanoic acid inhibitors of HIV-1 integrase and viral replication in cells, *J. Med. Chem.* *43*, 4923-4926.
18. Wai, J. S., Kim, B., Fisher, T. E., Zhuang, L., Embrey, M. W., Williams, P. D., Staas, D. D., Culberson, C., Lyle, T. A., Vacca, J. P., Hazuda, D. J., Felock, P. J., Schleif, W. A., Gabryelski, L. J., Jin, L., Chen, I. W., Ellis, J. D., Mallai, R., and Young, S. D. (2007) Dihydropyridopyrazine-1,6-dione HIV-1 integrase inhibitors, *Biorg. Med. Chem. Lett.* *17*, 5595-5599.
19. Hansch, C., Maloney, P. P., and Fujita, T. (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients, *Nature* *194*, 178 - 180.
20. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) 3-Dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature* *337*, 615-620.
21. Wlodawer, A. (2002) Rational approach to AIDS drug design through structural biology, *Annu. Rev. Med.* *53*, 595-614.
22. Ashburn, T. T., and Thor, K. B. (2004) Drug repositioning: Identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov.* *3*, 673-683.
23. Gills, J., Lo Piccolo, J., Tsurutani, J., Shoemaker, R. H., Best, C. J. M., Abu-Asab, M. S., Borojerdi, J., Warfel, N. A., Gardner, E. R., Danish, M., Hollander, M. C., Kawabata, S., Tsokos, M., Figga, W. D., Steeg, P. S., and Dennis, P. A. (2007) Nelfinavir, a lead HIV protease inhibitor, is a broad-spectrum, anticancer agent that induces endoplasmic reticulum stress, autophagy, and apoptosis in vitro and in vivo, *Clin. Cancer Res.* *13*, 5183-5194.
24. Shim, J. S., Rao, R., Beebe, K., Neckers, L., Han, I., Nahta, R., and Liu, J. O. (2012) Selective inhibition of HER2-positive breast cancer cells by the HIV protease Inhibitor nelfinavir, *J. Natl. Cancer Inst.* *104*, 1576-1590.
25. Martins, D., Beca, F., and Schmitt, F. (2014) Metastatic breast cancer: mechanisms and opportunities for cytology, *Cytopathology*.
26. DeSantis, C., Siegel, R., Bandi, P., and Jemal, A. (2011) Breast cancer statistics, 2011, *CA. Cancer J. Clin.* *61*, 409-418.
27. King, C., Kraus, M., and Aaronson, S. (1985) Amplification of a novel v-erbB-related gene in a human mammary carcinoma, *Sci* *229*, 974-976.
28. Slamon Dj Fau - Clark, G. M., Clark Gm Fau - Wong, S. G., Wong Sg Fau - Levin, W. J., Levin Wj Fau - Ullrich, A., Ullrich A Fau - McGuire, W. L., and

- McGuire, W. L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene, *Sci* 235, 177-182.
29. Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., Burchmore, M., Shak, S., Stewart, S. J., and Press, M. (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer, *J. Clin. Oncol.* 20, 719-726.
 30. Jane, T., Mehdi, M., Giuseppe, G., and Len, N. (2010) Targeting the dynamic HSP90 complex in cancer, *Nat. Rev. Cancer* 10, 537-549.
 31. Grover, A., Shandilya, A., Agrawal, V., Pratik, P., Bhasme, D., Bisaria, V. S., and Sundar, D. (2011) Hsp90/Cdc37 Chaperone/co-chaperone complex, a novel junction anticancer target elucidated by the mode of action of herbal drug Withaferin A, *BMC Bioinformatics* 12.
 32. Almerico, A. M., Tutone, M., Ippolito, M., and Lauria, A. (2007) Molecular modelling and QSAR in the discovery of HIV-1 integrase inhibitors, *Curr Comput Aided Drug Des* 3, 214-233.
 33. de Melo, E. B., and Castro Ferreira, M. M. (2009) Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* 44, 3577-3583.
 34. Dessalew, N. (2009) Investigation of the structural requirement for inhibiting HIV integrase: QSAR study, *Acta Pharmaceutica* 59, 31-43.
 35. Gupta, P., Roy, N., and Garg, P. (2009) Docking-based 3D-QSAR study of HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* 44, 4276-4287.
 36. Leonard, J. T., and Roy, K. (2008) Exploring molecular shape analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* 43, 81-92.
 37. Sharma, H., Cheng, X., and Buolamwini, J. K. (2012) Homology Model-Guided 3D-QSAR Studies of HIV-1 Integrase Inhibitors, *J. Chem. Inf. Model.* 52, 515-544.
 38. Srivastav, V. K., and Tiwari, M. (2013) QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors, *Arabian J. Chem. In press*, <http://www.sciencedirect.com/science/article/pii/S1878535213000270>
 39. Gupta, P., Garg, P., and Roy, N. (2012) Identification of Novel HIV-1 Integrase Inhibitors Using Shape-Based Screening, QSAR, and Docking Approach, *Chem Biol Drug Des* 79, 835-849.
 40. Cho, S. J., and Hermsmeier, M. A. (2002) Genetic algorithm guided selection: Variable selection and subset selection, *J. Chem. Inf. Comput. Sci.* 42, 927-936.
 41. Fernandez, M., Caballero, J., Fernandez, L., and Sarai, A. (2011) Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM), *Mol. Divers.* 15, 269-289.
 42. Hasegawa, K., Miyashita, Y., and Funatsu, K. (1997) GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists, *J. Chem. Inf. Comput. Sci.* 37, 306-310.
 43. Selwood, D. L., Livingstone, D. J., Comley, J. C. W., Odowd, A. B., Hudson, A. T., Jackson, P., Jandu, K. S., Rose, V. S., and Stables, J. N. (1990) Structure-

- activity-relationships of antifilarial antimycin analogs - a multivariate pattern-recognition study, *J. Med. Chem.* *33*, 136-142.
44. Cardozo, M. G., Imura, Y., Sugimoto, H., Yamanishi, Y., and Hopfinger, A. J. (1992) QSAR analyses of the substituted indanone and benzylpiperidine rings of a series of indanone benzylpiperidine inhibitors of acetylcholinesterase, *J. Med. Chem.* *35*, 584-589.
 45. Koehler, M. G., and Hopfinger, A. J. (1989) Molecular modelling of polymers: 5. Inclusion of intermolecular energetics in estimating glass and crystal-melt transition temperatures, *Polymer* *30*, 116-126.
 46. Goetz, M. P., Toft, D. O., Ames, M. M., and Erlichman, C. (2003) The Hsp90 chaperone complex as a novel target for cancer therapy, *Ann. Oncol.* *14*, 1169-1176.
 47. Peterson, L. B. (2012) Investigation of the Hsp90 C-terminal binding site, Novel inhibitors and isoform-dependent client proteins, In https://kuscholarworks.ku.edu/dspace/bitstream/1808/10218/1/Peterson_ku_0099_D_12224_DATA_1.pdf, 19 Oct. 2013.
 48. Arodola, O. A., Radhan Charan, D., and Mahmoud, E. S. S. (2014) QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor, *Lett Drug Des Discov* *11*, 618-627.

CHAPTER 2

Background on HIV/AIDS and HER2+ breast cancer

2.1. Introduction

This chapter briefly describes the background, life cycle and treatment of Human Immunodeficiency Virus (HIV), the causative agent of Acquired Immunodeficiency Syndrome (AIDS). This chapter focus on the inhibitors targeting HIV-1 integrase (IN), one of the key enzyme that is necessary for the completion of the HIV-1 life cycle and thereby a brief explanation about the repositioning of HIV-1 PIs for the treatment of HER2+ breast cancer inhibitors.

2.2. A brief history of HIV/AIDS

The earliest known case of a severe immune deficiency disease was reported among male homosexuals in 1981 and was characterized by infection related to suppressed immunity, unknown origin of fever and weight loss (1). These symptoms were reported by the Centre for Disease Control (CDC) in 1981 and christened as Acquired Immunodeficiency Syndrome (AIDS) (2). Few months later, the disease was reported to be transmissible through sexual contact (3), though the etiological agent remained a mystery. However, three years later, a group of researchers from France and the United States reported the causative agent of AIDS to be a lentivirus belonging to the retrovirus family (4, 5), which is the HIV virus (6, 7). There are two subtypes of HIV, HIV type 1 (HIV-1) and HIV type 2 (HIV-2) (8, 9). Numerous studies have been done on this retrovirus towards understanding its invasiveness, replication and pathogenicity (10-14).

The economic and social significance of HIV/AIDS is enormous and to date more than 60 million cases of HIV infections and approximately 30 million HIV-related deaths have been reported (15). Recent report shows that more than 40 million people are living with HIV (16), and of these 25% are on antiretroviral therapy (ART) (17). Sub-Saharan Africa is the most affected region, with nearly 1 in 20 adults living with HIV (18). In 2013, approximately 5.7 million people were living with HIV in South Africa (19).

2.3. The HIV-1 virus

The Human Immunodeficiency Virus type 1 (HIV-1 virus) belongs to the lentivirus (7), a subfamily of retroviruses (20). The HIV-1 virus, as shown in Figure 2.1, is composed of two identical single-stranded RNA, 2000 copies of viral proteins, three basic structural genes; group specific antigen (*gag*), polymerase (*pol*) and envelope (*env*) (21-23); and a few proteins from the host cell it infects, all surrounded by a lipid bilayer membrane. Additionally, two regulatory genes, trans-activator (*tat*) and regulator of expression of virion protein (*rev*), and four accessory genes, viral protein R (*vpr*), viral infectivity factor (*vif*), negative regulation factor (*nef*) and viral protein U (*vpu*) (viral protein X (*vpx*) in HIV-2) (24-30) help to regulate the life cycle of HIV-1 including translation, transcription, budding and assembly (31).

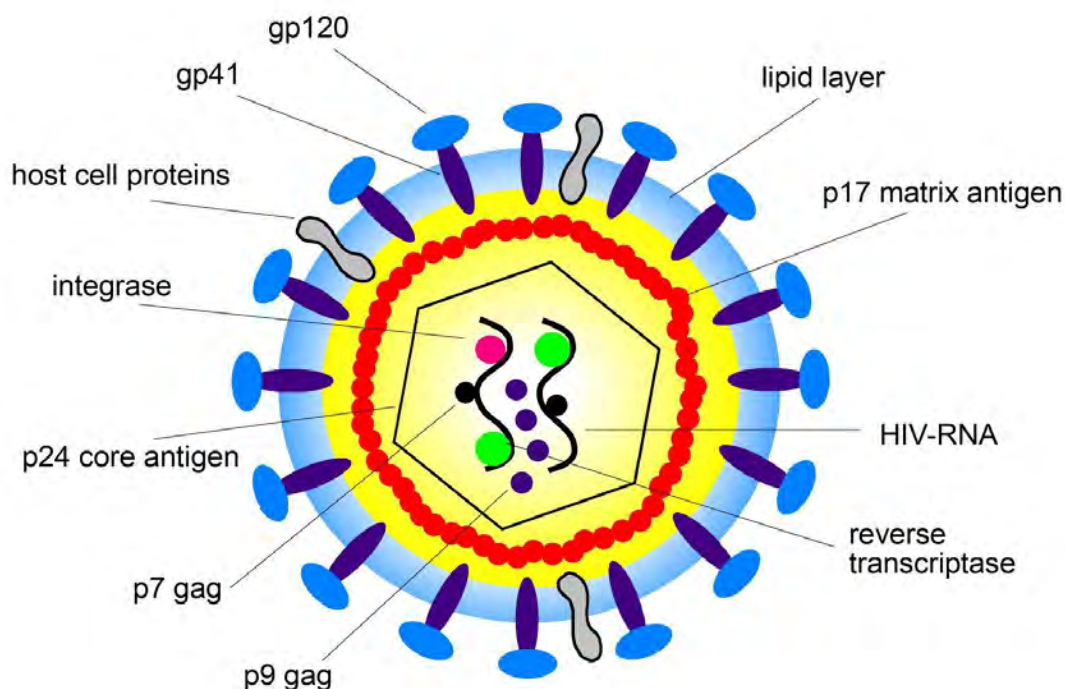


Figure 2.1. A model structure of the HIV-1 virus adapted from (32-34).

2.4. The HIV-1 life cycle

HIV-1 virus compromises the human immune system by targeting and destroying the CD4+ cells, which are a component of the T-lymphocytes. The life cycle of HIV-1 involves the train of events that can be grouped into six stages, namely: binding and fusion of the virus to the CD4 receptor on the host cell surface, reverse transcription, integration, transcription and translation, assembly and budding (35) [Figure 2.2].

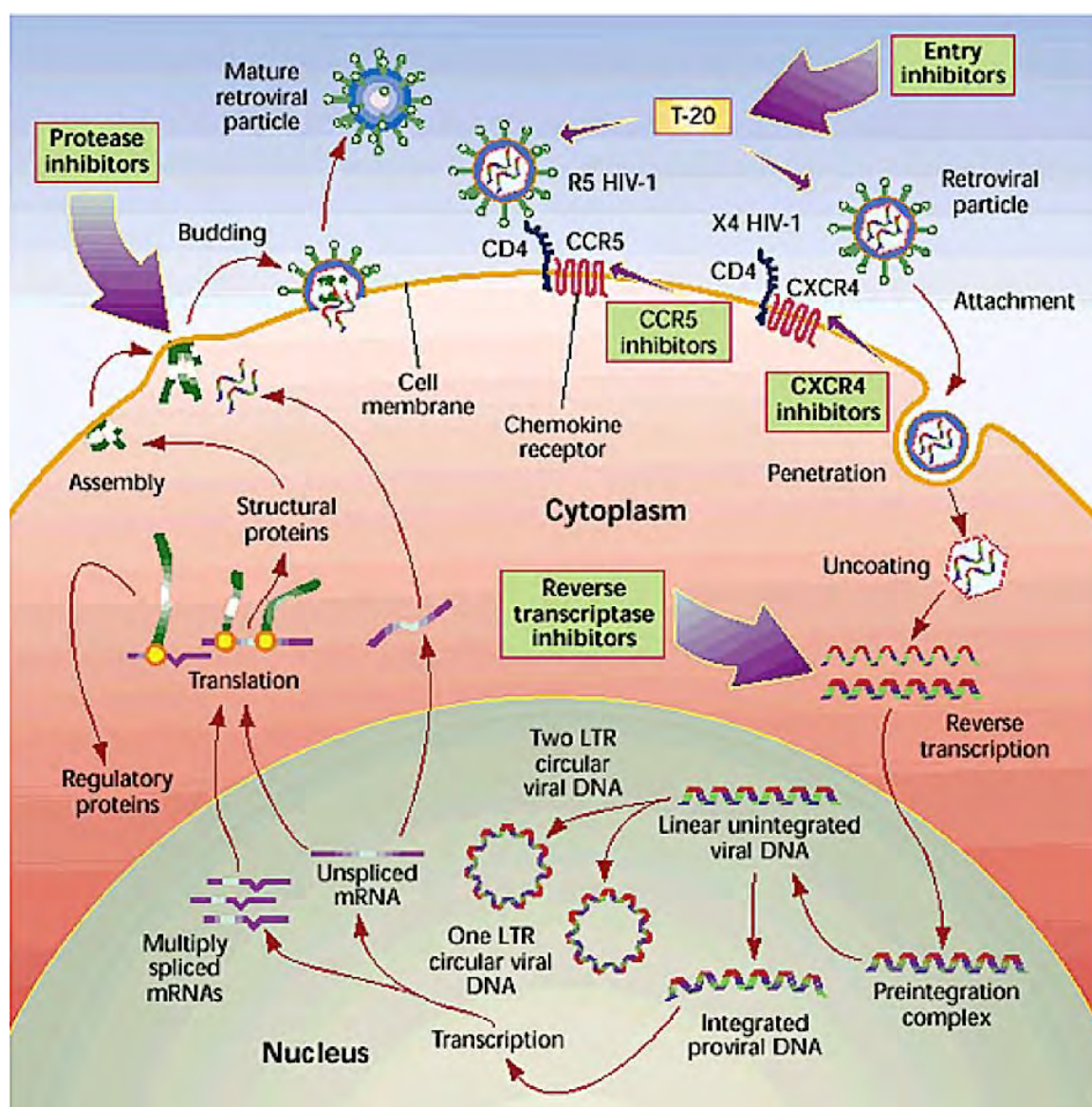


Figure 2.2. The replication cycle and targets for therapeutic intervention in the HIV life cycle (36).

The HIV-1 virus life cycle involves the following processes:

a. Binding and Fusion:

HIV begins its life cycle when the viral particle gains entry into vulnerable cells, which involves the interaction of CD4 molecules on the host cell surface with the HIV envelope glycoproteins gp120 and gp41 (37). Recently, the interaction of the virus with co-receptors on the cell surface, a crucial step in the entry stage, was found to be predominantly mediated by the chemokine receptors CCR5 and CXCR4 (38). The virus then fuses and releases its genetic material, RNA, into the host cell. The three major steps in viral entry have been targeted for drug development include inhibition of the CD4 binding, inhibition of co-receptor binding as well as the blocking of the gp41 conformational changes that permit viral fusion (39-42).

b. Reverse Transcription:

The viral enzyme, reverse transcriptase (RT) transcribes the single stranded RNA genome of the HIV virus to a double stranded DNA genome for integration into the host genome (43-45).

c. Integration:

The viral enzyme, integrase, mediates the integration of the double stranded HIV DNA into the host DNA by a process that involves 3'-processing and strand transfer reaction. The integrase splits the last two nucleotides from each 3'-end of the HIV DNA and then attaches them to the host DNA through a strand transfer reaction. The provirus (integrated HIV DNA) may remain idle for years, creating limited copies of the HIV-1 virus in this latent stage (46, 47).

d. Transcription and translation:

When the provirus is triggered to become active, the provirus recruits the host RNA polymerase to generate copies of the HIV genetic material, as well as the messenger RNA (mRNA). The mRNA is then translated to a long chain of polypeptide which is then cleaved to smaller viral proteins by the HIV protease (48). These cleaved proteins further undergo

posttranslational modification involving phosphorylation, methylation and glycosylation in order to become fully functional (49).

e. Assembly:

The viral proteins are then assembled together with copies of HIV's RNA genetic material to form new viral particles that is accumulated in the form of a bud (50).

f. Budding:

The newly accumulated virus pushes out 'buds' from the host cell. These buds steal part of the cell's outer envelope, which is covered with protein-sugar combinations called HIV glycoproteins. The glycoproteins are essential for the binding of the virus to CD4 and co-receptors. The new copies of HIV can then go on to infect other cells (49).

Consequently, each viral step in the HIV life cycle is a target for different inhibitors such as, the integrase strand transfer inhibitor (INSTI), non-nucleoside reverse transcriptase inhibitor (NNRTI), nucleoside reverse transcriptase inhibitor (NRTI) and protease inhibitor (PI). There are two basic classes of preventive measures against HIV-1 infection: One is targeting the post-entry stage, which is mostly used for treatment in antiretroviral therapy, and the other is targeting the entry stage, which can be used for both cure and preventive measures. In the next sub-section, the current post-entry stage treatment shall be discussed.

2.5. Current antiretroviral treatments for HIV-1 virus infection

The HIV-1 virus renders the body vulnerable to a variety of opportunistic infections, cancers, and other diseases (51-53). Currently, only drugs that lower the viral load exists and finding a cure for the disease has been a challenge (54). HIV-infected patients today, however, have a possible treatment option through the design of anti-HIV drugs. Current FDA approved HIV drugs interfere with the virus life cycle and they include protease inhibitors (PIs), reverse transcriptase (RT) inhibitors, integrase (IN) inhibitors and entry inhibitors (55). In 1987, zidovudine, a nucleoside reverse transcriptase inhibitor (NRTI) was approved by the FDA as the first antiretroviral drug for AIDS (56) and three years later, zalcitibine was approved for clinical use (57). In 1995, saquinavir was approved by FDA as the first protease inhibitor (58) and to date, there is a total of nine protease inhibitors. In 2003, the first fusion inhibitor,

enfuvirtide, was approved for clinical use (59). Some integrase inhibitors were also approved by the FDA; raltegravir (60), elvitegravir (61) and dolutegravir (62).

The next section discusses the inhibitors of the first three stages of the HIV-1 life cycle, more emphasis will be on the inhibitors on which this dissertation is based: integrase inhibitors and potential HER2+ breast cancer inhibitors.

2.5.1. Reverse transcriptase inhibitors

As HIV-1 is a retrovirus that replicates within a host cell, the RT enzyme is an essential component of HIV replication (43) and it is thus a major target for drug development (63). Presently, approved HIV-1 RT inhibitors used in antiretroviral therapy are grouped into two classes: (a) nucleoside analog RT inhibitors (NRTIs), which compete with naturally occurring nucleoside substrates for binding to the RT polymerase active site and after their incorporation into the primer site, they act as terminators of proviral DNA synthesis (64, 65), and (b) non-nucleoside RT inhibitors (NNRTIs), which bind to a hydrophobic pocket close to the RT active site where the allosteric inhibit RT enzymatic activity (64, 66). In 1987, zidovudine was the first NRTIs drug approved for HIV treatment and it is used alongside with other anti-HIV drugs such as lamuvidine and Abacavir as a combinant therapy. However, the increased resistance and side effects in zidovudine, such as nausea, myopathy, muscle tenderness, weakness to name a few, (67) lead to the approval of other drugs. Approved NNRTIs include that have been approved include nevirapine in 1996, efavirenz in 1998 and rilpivirine in 2011 (68).

2.5.2. Fusion inhibitors

This class of drug inhibits the binding, fusion and entry of the HIV-1 virus to the host cell. Enfuvirtide was approved in 2003 as a fusion inhibitor. Maraviroc, an entry inhibitor that inhibits the CCR5 co-receptor antagonist, was approved in 2007 (68).

2.5.3. Integrase inhibitors

HIV-1 integrase is a rational target for anti-HIV therapy. Drugs that inhibit integrase have been developed that include a set of compounds known as strand-transfer inhibitors (DKA and DKA-like inhibitors). These have been observed to be the most promising integrase

inhibitors. They can block strand transfer without affecting 3'-processing by chelating divalent co-factors in the integrase active site and by interfering with host DNA binding. Strand-transfer inhibitors, as candidate interfacial inhibitors, represent a new mechanism of action in drug discovery (69). These compounds share mechanistic and structural features (70-72), which will be reviewed in chapter 4. The first FDA-approved integrase inhibitor, raltegravir, was approved in 2007 (60), elvitegravir (61) approved in 2012 and dolutegravir (62) was recently approved in 2013 [Figure 2.3].

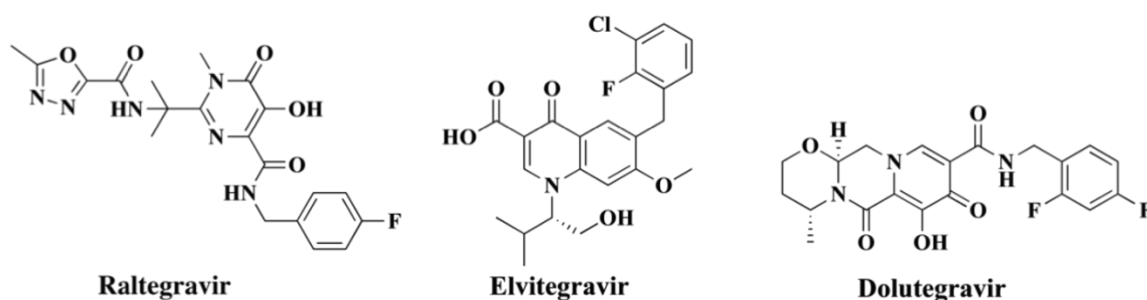


Figure 2.3. HIV-1 Integrase inhibitors (69).

2.5.4. Protease inhibitors

As was stated earlier in section 2.4, HIV-1 protease performs an important step in the life cycle of the HIV-1 virus where it cleaves the polypeptide into smaller viral proteins. Thus, it is an ideal target for HIV drug development. Protease inhibitor binds to the protease dimer interface, prevent the virus from maturing into its infectious stage, thus, blocking proliferation of the virus (73). Drugs, which target HIV-1 protease, remain one of the most remarkable achievements of medicine. The introduction of HIV protease inhibitors (PIs) in 1995 (58) and the application of highly active antiretroviral therapy (HAART), i.e., combination of PI with the HIV reverse transcriptase inhibitors, resulted in decreased mortality and a prolonged life expectancy of HIV-positive patients. Although the success of HIV-1 PIs has been remarkable and nine of these compounds are currently approved by the FDA as protease inhibitors [Figure 2.4], there is a need for the continued effort to develop more potent compounds (74, 75).

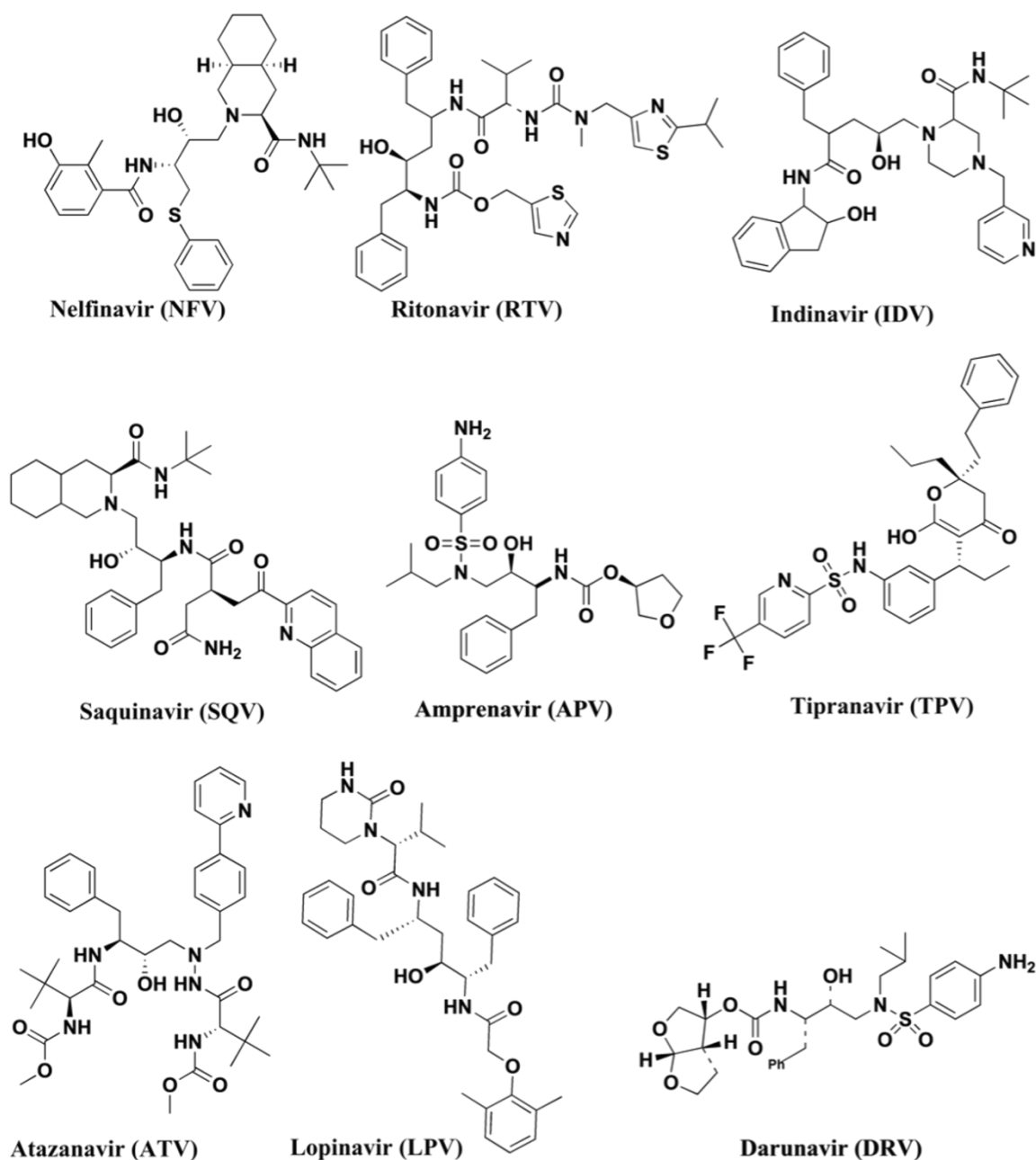


Figure 2.4. FDA-approved HIV-1 protease inhibitors (76).

The clinical practice of HIV-1 protease inhibitors (PIs) was brought into use in 1995 (77), when saquinavir (PDB code 1HXB) was approved for clinical trials as the first PR drug (58). The advantages of HIV-1 protease inhibitors include; prolonged viral control, better viral suppression, reduction of the indication of HIV-1 infection and reduction of death rates (77). HIV-1 PR treatment is therefore beneficial to HIV-1-infected patients as the mechanism of

action of the drug is to inhibit the catalytic activity of the protease, thus, preventing new virions from maturing into infectious forms (78).

2.6. HIV-1 viral enzymes

2.6.1. HIV-1 integrase enzyme

HIV-1 integrase is a 31 kDa protein, Y-shaped dimer that comprises of α and β chains [Figure 2.5]. The HIV-1 integrase comprises of three domains based on the vulnerability of the linker regions to proteolysis (79), functional studies (44, 79, 80), and the structures of the domains, which have been separately determined by x-ray crystallography or NMR. These includes an N-terminal domain which is the zinc binding site, a middle domain responsible for catalysis and a C-terminal DNA binding domain (80) [Figure 2.6].



Figure 2.5. The structure of the HIV-1 integrase enzyme (PDB code 1EX4) (32).

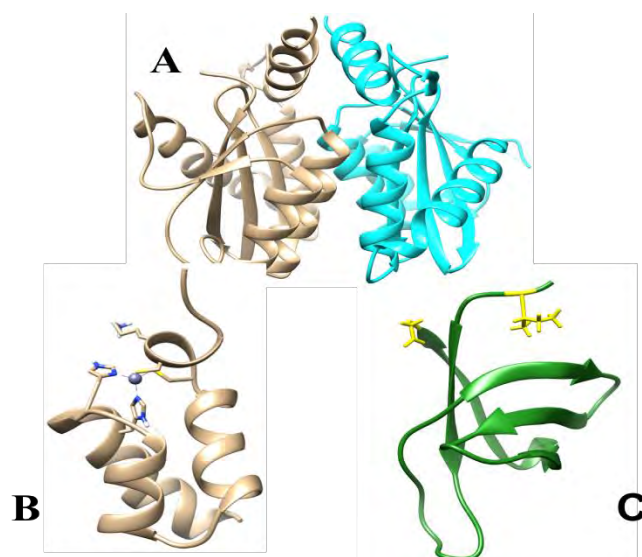


Figure 2.6. Structures of the three domains of HIV-1 integrase shown as ribbon diagrams (81). A, the catalytic core domain; B, the N-terminal domain; C, the C-terminal domain. PDB codes 1BIS, 1WJC, and 1IHV, respectively.

HIV-1, like other retroviruses, rely on the integration of its viral DNA into the host cell chromosomes so as to undergo the dormant stage without detection by the host immune system. In this way, HIV can remain latent in the cells for decades, thus, it can evade the immune system making its elimination from the body a challenge. This integration process is catalysed by HIV-1 integrase that performs a series of DNA cleaving and ligation reactions (82).

In the first step of integration is the 3'-end processing whereby two nucleotides are detached from each 3'-end processing. Cleavage occurs in the 3'-side of a CA dinucleotide; which is usually conserved among retroviruses, retro transposons and DNA transposons; thereby exposing the terminal 3'-hydroxyl group that would be joined to the host DNA. In the second step, the viral DNA end is inserted into the host DNA (83, 84) [Figure 2.7]. The sites of integration on the two host DNA strands are separated by 5 base pairs close to the integrated viral DNA; thus, two unpaired nucleotides at the 5'-ends of the viral DNA are removed and followed by ligation. The integrase is responsible for 3'-end processing and DNA strand transfer and other cellular enzymes carry out ligation (85, 86). As there is minimum specificity for the sites of integration in host DNA, insertion can take place at any location. Thus, HIV-1 IN has become a promising target for antiretroviral drug design (87-90) and new anti-HIV therapeutics (91, 92).

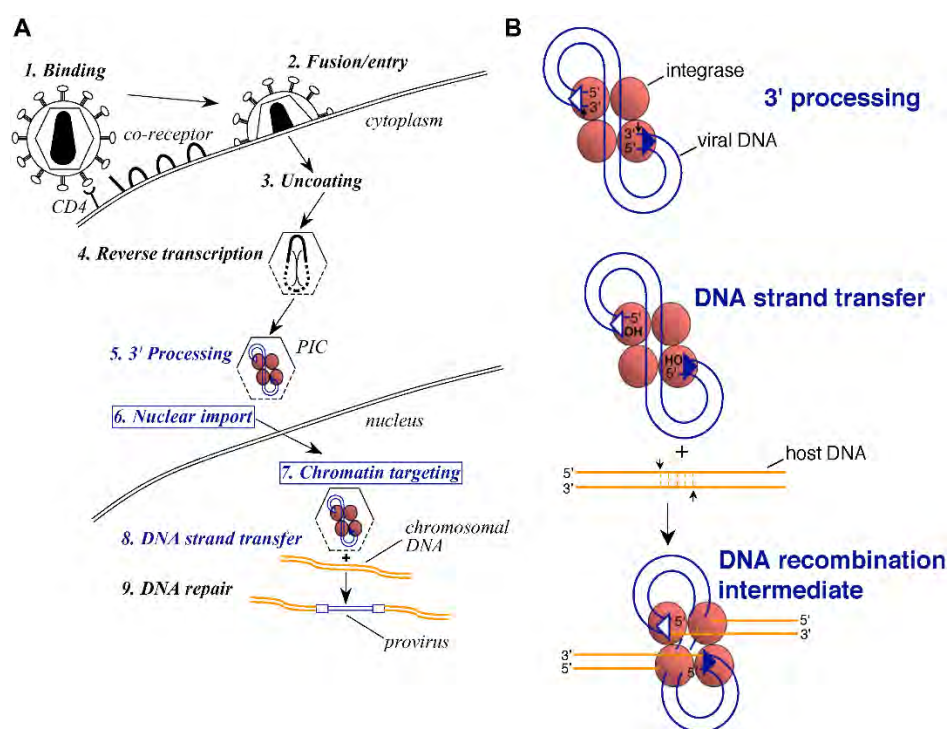


Figure 2.7. DNA cutting and joining steps in HIV integration taken from article (93).

2.6.2. HIV-1 protease enzyme

HIV-1 protease enzyme, an 11-kDa protein, 99 amino acids long, is classified as a C₂ symmetric active homodimeric aspartyl protease (77, 94). It is responsible for the cleavage of the newly synthesized viral *gag* and *gag-pol* poly-proteins to create mature and proper functionally active HIV virion (49). There are three domains in the enzyme shown in Figure 2.8 which includes the active site cavity (N-terminal domain), the middle domain and dimerization domain (C-terminal domain) (95). The enzyme contains a flexible flap region that closes down on the active site upon substrate binding (49). Active site cavity (N-terminal domain) plays an important role in the stabilization of the dimer and the catalytic site (49, 78, 94). Although the HIV-1 protease has been crystallized as a single monomer, it functions in the dimeric form. The dimer is observed in complexes of the HIV-1 protease with inhibitors (96, 97).

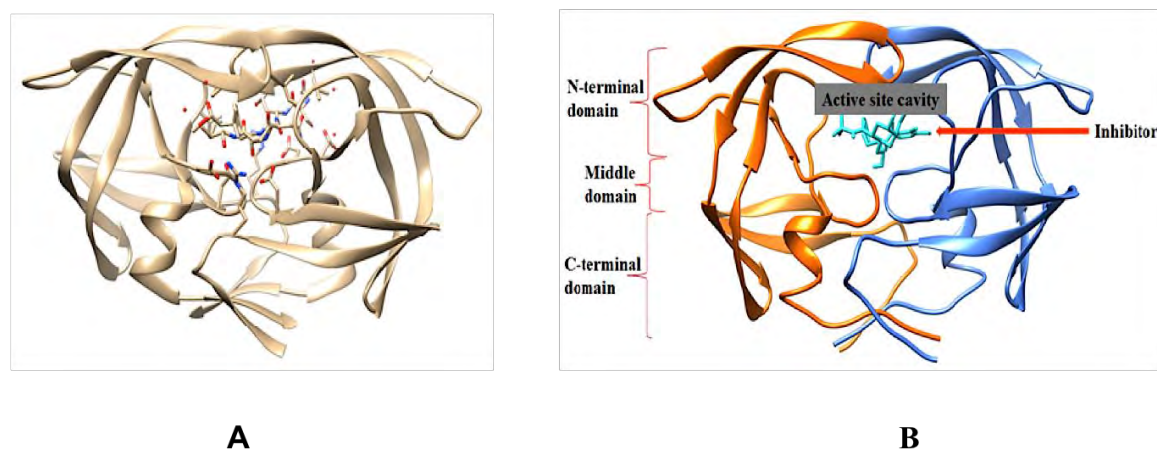


Figure 2.8. (A) Structure of HIV-1 protease in complex with an inhibitor (PDB code 7HVP) (98) taken from (99). (B) Structure of HIV-1 protease in complex with nelfinavir showing its active site (PDB code 1OHR) (100).

2.7. Brief Introduction to Drug Repositioning

During the past several years, there has been a surge of interest in drug repositioning. Biopharmaceutical industries have invested remarkable amounts in novel discovery technologies, such as structure-based drug design, combinatorial chemistry, high-throughput screening (HTS) and genomics, which were sold on the promise of improving productivity. However, there is still a gap in productivity, which has yielded few products. This productivity problem, alongside with global pressure on prices, challenges from ever-increasing regulatory hurdles, has obligated many drug developers to become more inventive in finding new uses for, and improved versions of, existing drugs. The process of finding new medical indications for existing drugs is known as repositioning. There are published reviews, which focused on repositioning and describe its general advantages over *de novo* drug discovery and development (101). A repositioned drug does not need the initial six to nine years typically required for the development of new drugs, but instead goes directly to preclinical testing and clinical trials, thus reducing risk and costs. Repositioning success stories are increasing in number. Example is sildenafil citrate (brand name: Viagra), which was repositioned from a common hypertension drug to a therapy for erectile dysfunction (102). Establishing accurate and efficient drug repositioning pipelines for specific studies requires the prioritization of existing computational methods based on the available knowledge or the development of new computational methods. Still, many challenges remain

for cost-effective drug repositioning studies. The complexities and richness of information available to drug repositioning studies largely determine their success rates (103).

2.8. The use of HIV-1 protease inhibitors as potential anticancer agents

It has also been hypothesized that “repositioning” of some HIV-1 protease inhibitors can inhibit HER2+ breast cancer. HER2+ breast cancer is considered to be caused by excessive production of HER2 gene (104). HER2 gene is important in the healthy growth and division of the breast cells, however, in 25-30% of breast cancers, the HER2 gene is overexpressed (105) as displayed in Figure 2.9.

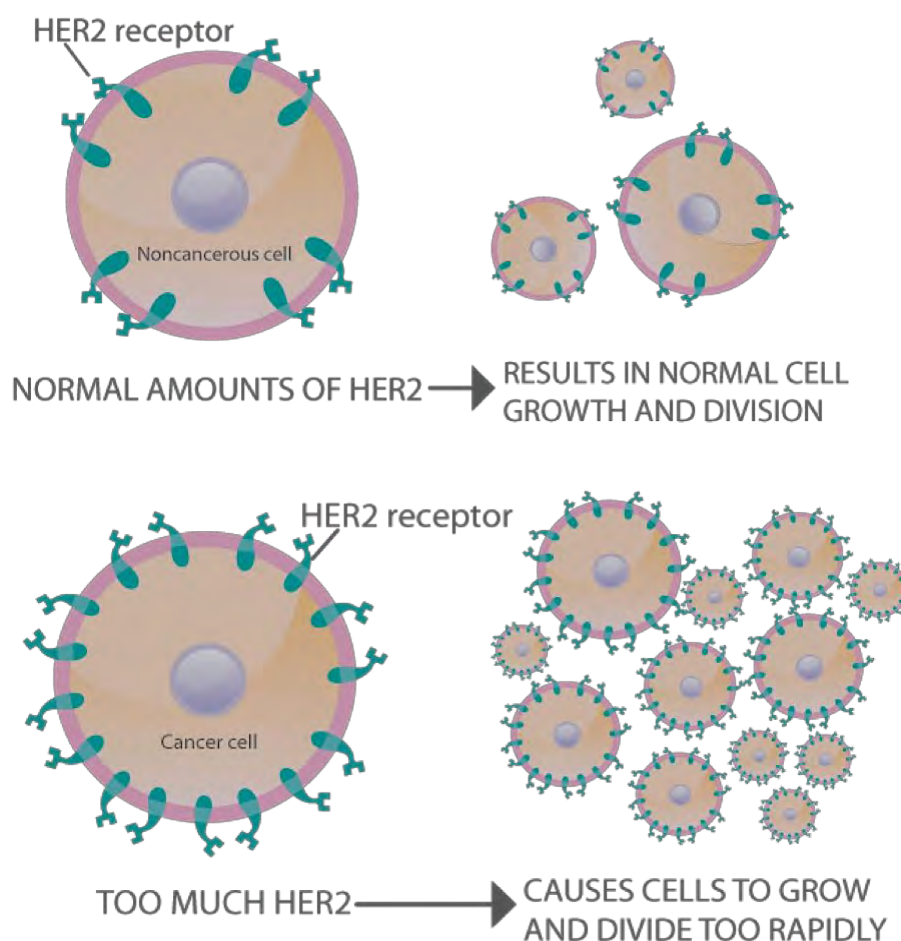


Figure 2.9. The model structure of a normal breast cell and cancer-infected breast cell retrieved from (106).

The Hsp90 has been reported to play a concise role in the overexpression of the HER2 gene (107). Therefore, to block HER2 overexpression, the necessary step would be to inhibit the

enzyme responsible for producing this estrogen, which is the Hsp90 (105, 108). The Hsp90 enzyme is responsible for protein folding (109-112), the proper functioning of about 200 client proteins (113, 114), to name a few. Hsp90 function as a homodimer and it has two important isoforms: the Hsp90 α and Hsp90 β (115-118). Each monomer of the Hsp90 dimer is comprised of four domains: a highly conserved N- and C-terminal domain, a middle domain and a charged linker that connects the N-terminal and middle domains (119-122) as shown in Figure 2.10. Chapter 5 holds detailed discussion about the Hsp90.

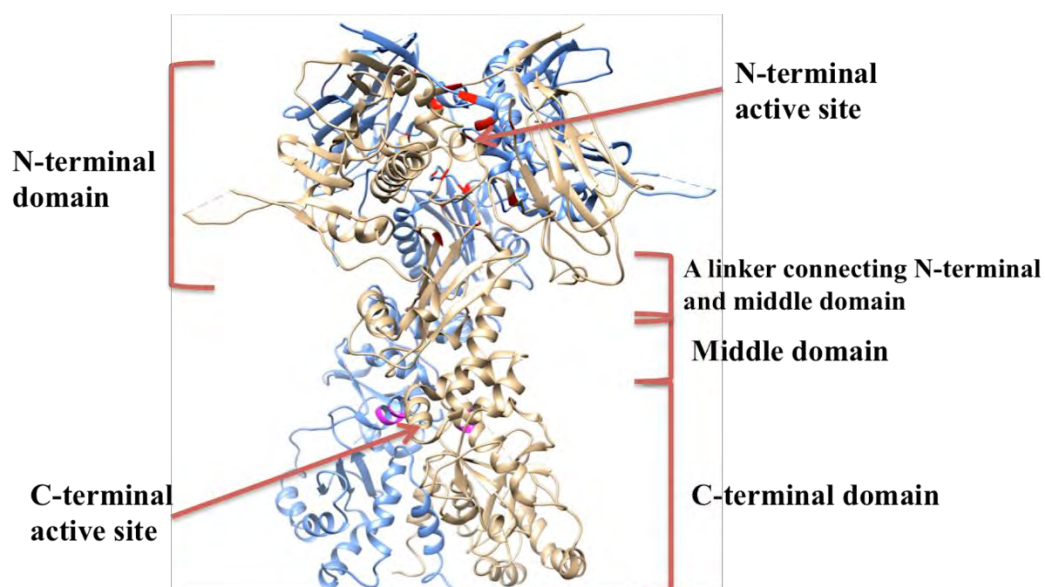


Figure 2.10. The crystal structure of Hsp90 Alpha (blue) and Beta chain (gold) (PDB code: 2CG9) showing its different domains (123).

Above all, the clinical evaluation of Hsp90 inhibitors has been met with some microbiological and biochemical challenges. Therapies that specifically target HER2 (124-127) are very effective and some has been shown to reduce breast cancer recurrence by as much as 40% (128). Despite this discovery, some HER2+ breast cancers do not respond to this treatment or they have become resistant to therapies (129).

Remarkably, it has recently been hypothesized that some anti-HIV drugs, specifically PIs, are dual inhibitors. This implies that not only can they inhibit HIV-1; they can also be used as anticancer agents through the process called “repositioning”. However, only few studies have been done in this regard (101, 130-132). The identification and characterization of these PIs

to be repositioned, form the basis of this study to gain an insight into the binding affinity of all recently approved FDA PIs against Hsp90, the enzyme responsible for initiating HER2+ breast cancer. This study explored the possibility of the Hsp90 (*N*- and *C*-terminal domain) binding affinity against nine protease inhibitors since it was also found that the CTD have some promising active site residues (*I33*) (details in chapter 5). Described in this thesis are the structure, design, and mechanistic features of novel HIV-1 PIs repositioned as potential human Hsp90 inhibitors that avoid the above-mentioned therapeutic liabilities of currently known Hsp90 inhibitors.

References:

1. Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981) Pneumocystis-carinii pneumonia and mucosal candidiasis in previously healthy homosexual men - evidence of a new acquired cellular immunodeficiency, *New Engl. J. Med.* 305, 1425-1431.
2. <http://healthfavo.com/hiv-virus-structure-anatomy-picture-reference.html>.
3. Masur, H., Michelis, M. A., Wormser, G. P., Lewin, S., Gold, J., Tapper, M. L., Giron, J., Lerner, C. W., Armstrong, D., Setia, U., Sender, J. A., Siebken, R. S., Nicholas, P., Arlen, Z., Maayan, S., Ernst, J. A., Siegal, F. P., and Cunninghamrundles, S. (1982) Opportunistic infection in previously healthy women - initial manifestations of a community-acquired cellular immunodeficiency, *Ann. Intern. Med.* 97, 533-539.
4. Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P. D. (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-iii) from patients with AIDS and at risk for AIDS, *Sci* 224, 500-503.
5. Barresinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Daugey, C., Axlerblin, C., Vezinetbrun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune-deficiency syndrome (AIDS), *Sci* 220, 868-871.
6. Gallo, R. C., and Montagnier, L. (1988) AIDS in 1988, *SciAm* 259, 41-48.
7. Popovic M Fau - Sarngadharan, M. G., Sarngadharan Mg Fau - Read, E., Read E Fau - Gallo, R. C., and Gallo, R. C. (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS, *Sci* 224, 497-500.
8. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M., Santos-Ferreira, M., Laurent, A., Daugey, C., Katlama, C., and Rouzioux, C. (1986) Isolation of a new human retrovirus from West African patients with AIDS, *Sci* 233, 343-346.
9. Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L., and Alizon, M. (1986) Molecular-cloning and polymorphism of the human immune-deficiency virus type-2, *Nature* 324, 691-695.
10. Perno, C. F., Yarchoan, R., Cooney, D. A., Hartman, N. R., Webb, D. S. A., Hao, Z., Mitsuya, H., Johns, D. G., and Broder, S. (1989) Replication of Human Immunodeficiency Virus in monocytes - granulocyte macrophage colony-stimulating factor (GM-CSF) potentiates viral production yet enhances the antiviral effect mediated by 3'-azido-2'3'-dideoxythymidine (AZT) and other dideoxynucleoside congeners of thymidine, *J. Exp. Med.* 169, 933-951.
11. Bergamaschi, A., and Pancino, G. (2010) Host hindrance to HIV-1 replication in monocytes and macrophages, *Retrovirology* 7.
12. Wain, L. V., Bailes, E., Bibollet-Ruche, F., Decker, J. M., Keele, B. F., Van Heuverswyn, F., Li, Y., Takehisa, J., Ngole, E. M., Shaw, G. M., Peeters, M., Hahn, B. H., and Sharp, P. M. (2007) Adaptation of HIV-1 to its human host, *Mol. Biol. Evol.* 24, 1853-1860.
13. Li, G., Bukrinsky, M., and Zhao, R. Y. (2009) HIV-1 Viral Protein R (VPR) and its Interactions with Host Cell, *Curr. HIV Res.* 7, 178-183.
14. Hladik, F., and McElrath, M. J. (2008) Setting the stage: host invasion by HIV, *Nat. Rev. Immunol.* 8, 447-457.

15. Merson, M. H., O'Malley J Fau - Serwadda, D., Serwadda D Fau - Apisuk, C., and Apisuk, C. (2008) The history and challenge of HIV prevention, *Lancet Infect. Dis.* 372, 497-500.
16. UNAIDS. (2011) UNAIDS world's AID day report, http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/jc2216_worlddaidsday_report_2011_en.pdf.
17. UNAIDS. (http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf).
18. UNAIDS. (http://data.unaids.org/pub/GlobalReport/2008/jc1510_2008_global_report_pp211_234_en.pdf) 20 Jan 2014.
19. UNAIDS. (http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/201309_epi_core_en.pdf) 20 Jan 2014.
20. Sharp, P. M., and Hahn, B. H. (2011) Origins of HIV and the AIDS Pandemic, *Cold Spring Harb. Perspect. Med.* 1.
21. Checkley, M. A., Lutge, B. G., and Freed, E. O. (2011) HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation, *JMBio* 410, 582-608.
22. Hill, M., Tachedjian, G., and Mak, J. (2005) The packaging and maturation of the HIV-1 pol proteins, *Curr. HIV Res.* 3, 73-85.
23. Klein, K. C., Reed, J. C., and Lingappa, J. R. (2007) Intracellular destinies: Degradation, targeting, assembly, and endocytosis of HIV gag, *Aids Reviews* 9, 150-161.
24. Grewe, B., and Ueberla, K. (2010) The human immunodeficiency virus type 1 Rev protein: menage a trois during the early phase of the lentiviral replication cycle, *J. Gen. Virol.* 91, 1893-1897.
25. Laguette, N., Bregnard, C., Benichou, S., and Basmaciogullari, S. (2010) Human immunodeficiency virus (HIV) type-1, HIV-2 and simian immunodeficiency virus Nef proteins, *Mol. Aspects Med.* 31, 418-433.
26. Laguette, N., Sobhian, B., Casartelli, N., Ringiard, M., Chable-Bessia, C., Segeral, E., Yatim, A., Emiliani, S., Schwartz, O., and Benkirane, M. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx, *Nature* 474, 654-U132.
27. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts, *Nature* 424, 99-103.
28. Neil, S. J. D., Zang, T., and Bieniasz, P. D. (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu, *Nature* 451, 425-430.
29. Planelles, V., and Benichou, S. (2009) Vpr and Its Interactions with Cellular Proteins, *Curr. Top. Microbiol. Immunol.* 339, 177-200.
30. Romani, B., Engelbrecht, S., and Glashoff, R. H. (2010) Functions of Tat: the versatile protein of human immunodeficiency virus type 1, *J. Gen. Virol.* 91, 1-12.
31. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV, *JMBio* 285, 1-32.
32. Chen, J. C. H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding, *Proc. Natl. Acad. Sci. U. S. A.* 97, 8233-8238.

33. Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M., and Vogt, V. (1988) Standardized and simplified nomenclature for proteins common to all retroviruses, *J. Virol.* *62*, 1808-1809.
34. Welker, R., Hohenberg, H., Tessmer, U., Huckhagel, C., and Krausslich, H. G. (2000) Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1, *J. Virol.* *74*, 1168-1177.
35. Moore, J. P., and Stevenson, M. (2000) New targets for inhibitors of HIV-1 replication, *Nat. Rev. Mol. Cell Biol* *1*, 40-49.
36. Michael, N. L., and Moore, J. P. (1999) HIV-1 entry inhibitors: Evading the issue, *Nat. Med.* *5*, 740-742.
37. Jones, P. L. S., Korte, T., and Blumenthal, R. (1998) Conformational changes in cell surface HIV-1 envelope glycoproteins are triggered by cooperation between cell surface CD4 and co-receptors, *J. Biol. Chem.* *273*, 404-409.
38. Cilliers, T., Nhlapo, J., Coetzer, M., Orlovic, D., Ketas, T., Olson, W. C., Moore, J. P., Trkola, A., and Morris, L. (2003) The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C, *J. Virol.* *77*, 4449-4456.
39. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease, *Annu. Rev. Immunol.* *17*, 657-700.
40. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature* *393*, 648-659.
41. D'Souza, M. P., Cairns, J. S., and Plaeger, S. F. (2000) Current evidence and future directions for targeting HIV entry - therapeutic and prophylactic strategies, *J. Am. Med. Assoc.* *284*, 215-222.
42. Strizki, J. M., Xu, S., Wagner, N. E., Wojcik, L., Liu, J., Hou, Y., Endres, M., Palani, A., Shapiro, S., Clader, J. W., Greenlee, W. J., Tagat, J. R., McCombie, S., Cox, K., Fawzi, A. B., Chou, C. C., Pugliese-Sivo, C., Davies, L., Moreno, M. E., Ho, D. D., Trkola, A., Stoddart, C. A., Moore, J. P., Reyes, G. R., and Baroudy, B. M. (2001) SCH-C (SCH 351125), an orally bioavailable, small molecule antagonist of the chemokine receptor CCR5, is a potent inhibitor of HIV-1 infection in vitro and in vivo, *Proc. Natl. Acad. Sci. U. S. A.* *98*, 12718-12723.
43. Esposito, F., Corona, A., and Tramontano, E. (2012) HIV-1 reverse transcriptase still remains a new drug target: structure, function, classical inhibitors, and new inhibitors with innovative mechanisms of actions, *Mol. Biol. Int.* *2012*, 1-23.
44. Vangent, D. C., Vink, C., Groeneger, A., and Plasterk, R. H. A. (1993) Complementation between HIV integrase proteins mutated in different domains, *EMBO J.* *12*, 3261-3267.
45. Tronchet, J. M. J., and Seman, M. (2003) Nonnucleoside inhibitors of HIV-1 reverse transcriptase: From the biology of reverse transcription to molecular design, *Curr. Top. Med. Chem.* *3*, 1496-1511.
46. Declercq, E. (1995) Toward improved anti-HIV chemotherapy-therapeutic strategies for intervention with HIV-infections, *J. Med. Chem.* *38*, 2491-2517.
47. Vandegraaff, N., and Engelman, A. (2007) Molecular mechanisms of HIV integration and therapeutic intervention, *Expert Rev. Mol. Med.* *9*, 1-19.
48. Van Lint, C., Bouchat, S., and Marcello, A. (2013) HIV-1 transcription and latency: an update, *Retrovirology* *10*.

49. Brik, A., and Wong, C. H. (2003) HIV-1 protease: mechanism and drug discovery, *Org. Biomol. Chem.* 1, 5-14.
50. Spearman, P. (2006) Cellular cofactors involved in HIV assembly and budding, *Curr Opin HIV AIDS* 1, 200-207.
51. Gait, M. J., and Karn, J. (1995) Progress in anti-HIV structure-based drug design, *Trends Biotechnol.* 13, 430-438.
52. Andrade, M. D., and Skalka, A. M. (1996) Retroviral integrase, putting the pieces together, *J. Biol. Chem.* 271, 19633-19636.
53. Nanni, R. G., Ding, J., Jacobo-Molina, A., Hughes, S. H., and Arnold, E. (1993) Review of HIV-1 reverse transcriptase three-dimensional structure: implications for drug design, *Perspect. Drug Discov. Des.* 1, 129-150.
54. Cohen, M. S., Smith, M. K., Muessig, K. E., Hallett, T. B., Powers, K. A., and Kashuba, A. D. (2013) Antiretroviral treatment of HIV-1 prevents transmission of HIV-1: where do we go from here?, *The Lancet* 382, 1515-1524.
55. Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Curr. Drug Targets: Infect. Disord.* 2, 17-32.
56. Marwick, C. (1987) AZT (zidovudine) just a step away from fda approval for aids therapy, *JAMA* 257, 1281-1282.
57. Anastasi, J. K., and Rivera, J. L. (1991) AIDS drug update: DDI and DDC, *RN* 54, 41-43.
58. James, J. S. (1995) Saquinavir (Invirase): first protease inhibitor approved-reimbursement, information hotline numbers, *AIDS Treat News*, 1-2.
59. Williams, I. G. (2003) Enfuvirtide (Fuzeon (R)): The first fusion inhibitor, *Int. J. Clin. Pract.* 57, 890-897.
60. Hicks, C., and Gulick, R. M. (2009) Raltegravir: The first HIV type 1 integrase inhibitor, *Clin. Infect. Dis.* 48, 931-939.
61. FDA. (2012) NDA 203100, <http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/antiviraldrugsadvisorycommittee/ucm305852.pdf>.
62. FDA. Antiretroviral drugs used in the treatment of HIV infection, <http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm>.
63. Iliina, T., LaBarge, K., Sarafianos, S. G., Ishima, R., and Parniak, M. A. (2012) Inhibitors of HIV-1 reverse transcriptase-associated ribonuclease H activity, *Biology* 1, 521-541.
64. Vivet-Boudou, V., Didierjean, J., Isel, C., and Marquet, R. (2006) Nucleoside and nucleotide inhibitors of HIV-1 replication, *Cell. Mol. Life Sci.* 63, 163-186.
65. Bauman, J. D., Das, K., Ho, W. C., Baweja, M., Himmel, D. M., Clark, A. D., Oren, D. A., Boyer, P. L., Hughes, S. H., Shatkin, A. J., and Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design, *NAR* 36, 5083-5092.
66. Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses *Antiviral Res.* 72, 256-256.
67. Max, B., and Sherer, R. (2000) Management of the adverse effects of antiretroviral therapy and medication adherence, *Clin. Infect. Dis.* 30, S96-S116.
68. FDA HIV approved drugs. (<http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm>).

69. Pommier, Y., Johnson, A. A., and Marchand, C. (2005) Integrase inhibitors to treat HIV/AIDS, *Nat. Rev. Drug Discov.* 4, 236-248.
70. Carreau, S., Mouscadet, J. F., Goulaouic, H., Subra, F., and Auclair, C. (1993) Inhibitory effect of the polyanionic drug suramin on the in-vitro HIV DNA integration reaction, *Arch. Biochem. Biophys.* 305, 606-610.
71. Cushman, M., and Sherman, P. (1992) Inhibition of HIV-1 integration protein by aurintricarboxylic acid monomer analogs, and polymer fractions, *BBRC* 185, 85-90.
72. Fesen, M. R., Kohn, K. W., Leteurtre, F., and Pommier, Y. (1993) Inhibitors of human immunodeficiency virus integrase, *Proc. Natl. Acad. Sci. U. S. A.* 90, 2399-2403.
73. Heal, J. W., Jimenez-Roldan, J. E., Wells, S. A., Freedman, R. B., and Roemer, R. A. (2012) Inhibition of HIV-1 protease: the rigidity perspective, *Bioinformatics* 28, 350-357.
74. Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J., and Hatada, M. H. (1991) Novel binding of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere, *J. Med. Chem.* 34, 3340-3342.
75. Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., and Navia, M. A. (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme, *J. Am. Chem. Soc.* 117, 1181-1182.
76. Hughes, P. J., Cretton-Scott, E., Teague, A., and Wensel, T. M. (2011) Protease Inhibitors for Patients With HIV-1 Infection: A Comparative Overview, *P T.* 36, 332-345.
77. Martinez-Cajas, J. L., and Wainberg, M. A. (2007) Protease inhibitor resistance in HIV-infected patients: Molecular and clinical perspectives, *Antiviral Res.* 76, 203-221.
78. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annu. Rev. Biophys. Biomol. Struct.* 27, 249-284.
79. Engelman, A., and Craigie, R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro, *J. Virol.* 66, 6361-6369.
80. Engelman, A., Bushman, F. D., and Craigie, R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex, *EMBO J.* 12, 3269-3275.
81. Carson, M. (1987) Ribbon models of macromolecules, *J. Mol. Graphics* 5, 103-106.
82. Chiu, T. K., and Davies, D. R. (2004) Structure and function of HIV-1 integrase, *Curr. Top. Med. Chem.* 4, 965-977.
83. Craigie, R. (2001) HIV integrase, a brief overview from chemistry to therapeutics, *J. Biol. Chem.* 276, 23213-23216.
84. Di Santo, R., Costi, R., Artico, M., Tramontano, E., La Colla, P., and Pani, A. (2003) HIV-1 integrase inhibitors that block HIV-1 replication in infected cells. Planning synthetic derivatives from natural products, *Pure Appl. Chem.* 75, 195-206.
85. Brin, E., Yi, J., Skalka, A. M., and Leis, J. (2000) Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration, *J. Biol. Chem.* 275, 39287-39295.
86. Yoder, K. E., and Bushman, F. D. (2000) Repair of gaps in retroviral DNA integration intermediates, *J. Virol.* 74, 11191-11200.
87. Barreca, M. L., Ferro, S., Rao, A., De Luca, L., Zappala, M., Monforte, A. M., Debyser, Z., Witvrouw, M., and Chimirri, A. (2005) Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors, *J. Med. Chem.* 48, 7084-7088.

88. d'Angelo, J., Mouscadet, J. F., Desmaële, D., Zouhiri, F., and Leh, H. (2001) HIV-1 integrase: the next target for AIDS therapy?, *Pathol. Biol.* 49, 237-246.
89. Grobler, J. A., Stillmock, K., Hu, B. H., Witmer, M., Felock, P., Espeseth, A. S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J. S., Young, S., Vacca, J., and Hazuda, D. J. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes, *Proc. Natl. Acad. Sci. U. S. A.* 99, 6661-6666.
90. Zhuang, L. C., Wai, J. S., Embrey, M. W., Fisher, T. E., Egbertson, M. S., Payne, L. S., Guare, J. P., Vacca, J. P., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Witmer, M. V., Moyer, G., Schleif, W. A., Gabryelski, L. J., Leonard, Y. M., Lynch, J. J., Michelson, S. R., and Young, S. D. (2003) Design and synthesis of 8-hydroxy- 1,6 naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells, *J. Med. Chem.* 46, 453-456.
91. Englund, G., Theodore, T. S., Freed, E. O., Engelman, A., and Martin, M. A. (1995) Integration is required for productive infection of monocyte-derived macrophages by human immunodeficiency virus type 1, *J. Virol.* 69, 3216-3219.
92. Bushman, F. D., and Craigie, R. (1991) Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA, *Proc. Natl. Acad. Sci. U. S. A.* 88, 1339-1343.
93. Engelman, A., and Cherepanov, P. (2008) The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication, *PLoS Path.* 4, e1000046.
94. Castro, H. C., Abreu, P. A., Geraldo, R. B., Martins, R. C. A., dos Santos, R., Loureiro, N. I. V., Cabral, L. M., and Rodrigues, C. R. (2011) Looking at the proteases from a simple perspective, *J. Mol. Recognit.* 24, 165-181.
95. Weber, I. T., and Agniswamy, J. (2009) HIV-1 protease: Structural perspectives on drug resistance, *Viruses-Basel* 1, 1110-1136.
96. Spinelli, S., Liu, Q. Z., Alzari, P. M., Hirel, P. H., and Poljak, R. J. (1991) The 3-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU, *Biochimie* 73, 1391-1396.
97. Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E. M., Sayed, Y., Ibrahim, M. A. A., Naicker, P., and Soliman, M. E. S. (2013) Comparison of the molecular dynamics and calculated binding free energies for nine FDA-approved HIV-1 PR drugs against subtype B and C-SA HIV PR, *Chem. Biol. Drug Des.* 81, 208-218.
98. Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S. B. H., and Wlodawer, A. (1990) X-RAY crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus-1 and a substrate-based hydroxyethylamine inhibitor, *Proc. Natl. Acad. Sci. U. S. A.* 87, 8805-8809.
99. Sriphanlop, P. (2009) HIV-1 protease: A target for AIDS therapy, http://people.mbi.ucla.edu/yeates/153AH_2009_project/sriphanlop.html 04 Jan 2014.
100. Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997) Viracept (nelfinavir mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease, *J. Med. Chem.* 40, 3979-3985.
101. Ashburn, T. T., and Thor, K. B. (2004) Drug repositioning: Identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov.* 3, 673-683.
102. Novac, N. (2013) Challenges and opportunities of drug repositioning, *Trends Pharmacol. Sci.* 34, 267-272.

103. Jin, G., and Wong, S. T. C. (2014) Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines, *Drug Discov. Today* 19, 637-644.
104. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) Human-breast cancer - correlation of relapse and survival with amplification of the HER-2 neu oncogene, *Sci* 235, 177-182.
105. Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., Burchmore, M., Shak, S., Stewart, S. J., and Press, M. (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer, *J. Clin. Oncol.* 20, 719-726.
106. Burstein, H. J. (2005) The distinctive nature of HER2-positive breast cancers, *New Engl. J. Med.* 353, 1652-1654.
107. Kurebayashi, J. (2001) Biological and clinical significance of HER2 overexpression in breast cancer, *Breast Cancer* 8, 45-51.
108. Grover, A., Shandilya, A., Agrawal, V., Pratik, P., Bhasme, D., Bisaria, V. S., and Sundar, D. (2011) Hsp90/Cdc37 Chaperone/co-chaperone complex, a novel junction anticancer target elucidated by the mode of action of herbal drug Withaferin A, *BMC Bioinformatics* 12.
109. Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S. M., Panaretou, B., Piper, P. W., and Pearl, L. H. (2003) Structural and functional analysis of the middle segment of Hsp90: Implications for ATP hydrolysis and client protein and cochaperone interactions, *Mol. Cell* 11, 647-658.
110. Pratt, W. B., and Toft, D. O. (2003) Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery, *Exp. Biol. Med.* 228, 111-133.
111. Soti, C., Nagy, E., Giricz, Z., Vigh, L., Csermely, P., and Ferdinandy, P. (2005) Heat shock proteins as emerging therapeutic targets, *Br. J. Pharmacol.* 146, 769-780.
112. Chiosis, G., Vilenchik, M., Kim, J., and Solit, D. (2004) Hsp90: the vulnerable chaperone, *Drug Discov. Today* 9, 881-888.
113. Picard, D. (2014) Hsp90 interactors, <http://www.picard.ch/downloads/Hsp90interactors.pdf>.
114. Picard, D. (2014) Hsp90 facts & literature, <http://www.picard.ch/downloads/Hsp90facts.pdf>.
115. Chen, B., Piel, W. H., Gui, L. M., Bruford, E., and Monteiro, A. (2005) The HSP90 family of genes in the human genome: Insights into their divergence and evolution, *Genomics* 86, 627-637.
116. Johnson, J. L. (2012) Evolution and function of diverse Hsp90 homologs and cochaperone proteins, *Biochim. Biophys. Acta* 1823, 607-613.
117. Csermely, P., Schnaider, T., Soti, C., Prohaszka, Z., and Nardai, G. (1998) The 90-kDa molecular chaperone family: Structure, function, and clinical applications. A comprehensive review, *Pharmacol. Ther.* 79, 129-168.
118. Sreedhar, A. S., Kalmar, E., Csermely, P., and Shen, Y. F. (2004) Hsp90 isoforms: functions, expression and clinical importance, *FEBS Lett.* 562, 11-15.
119. Ali, M. M. U., Roe, S. M., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C., and Pearl, L. H. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex, *Nature* 440, 1013-1017.
120. Krishna, P., and Gloor, G. (2001) The Hsp90 family of proteins in Arabidopsis thaliana, *Cell Stress Chaperones* 6, 238-246.
121. Nemoto, T., Sato, N., Iwanari, H., Yamashita, H., and Takagi, T. (1997) Domain structures and immunogenic regions of the 90-kDa heat-shock protein (HSP90) -

- Probing with a library of anti-HSP90 monoclonal antibodies and limited proteolysis, *J. Biol. Chem.* 272, 26179-26187.
122. Prodromou, C., and Pearl, L. H. (2003) Structure and functional relationships of Hsp90, *Curr. Cancer Drug Targets* 3, 301-323.
 123. Dollins, D. E., Warren, J. J., Immormino, R. M., and Gewirth, D. T. (2007) Structures of GRP94-Nucleotide complexes reveal mechanistic differences between the hsp90 chaperones, *Mol. Cell* 28, 41-56.
 124. Eiermann, W. (2001) Trastuzumab combined with chemotherapy for the treatment of HER2-positive metastatic breast cancer: Pivotal trial data, *Ann. Oncol.* 12, 57-62.
 125. Keating, G. M. (2012) Pertuzumab In the First-Line Treatment of HER2-Positive Metastatic Breast Cancer, *Drugs* 72, 353-360.
 126. Sendur, M. A. N., Aksoy, S., and Altundag, K. (2012) Pertuzumab in HER2-positive breast cancer, *Curr. Med. Res. Opin.* 28, 1709-1716.
 127. Collins, D., Hill, A. D. K., and Young, L. (2009) Lapatinib: A competitor or companion to trastuzumab?, *Cancer Treat. Rev.* 35, 574-581.
 128. McKeage, K., and Perry, C. M. (2002) Trastuzumab - A review of its use in the treatment of metastatic breast cancer overexpressing HER2, *Drugs* 62, 209-243.
 129. Pohlmann, P. R., Mayer, I. A., and Mernaugh, R. (2009) Resistance to Trastuzumab in Breast Cancer, *Clin. Cancer. Res.* 15, 7479-7491.
 130. Fuentes, G., Scaltriti, M., Baselga, J., and Verma, C. S. (2011) Synergy between trastuzumab and pertuzumab for human epidermal growth factor 2 (Her2) from colocalization: an in silico based mechanism, *Breast Cancer Res.* 13.
 131. Shim, J. S., Rao, R., Beebe, K., Neckers, L., Han, I., Nahta, R., and Liu, J. O. (2012) Selective Inhibition of HER2-Positive Breast Cancer Cells by the HIV Protease Inhibitor Nelfinavir, *J Natl. Cancer Inst.* 104, 1576-1590.
 132. Gills, J., Lo Piccolo, J., Tsurutani, J., Shoemaker, R. H., Best, C. J. M., Abu-Asab, M. S., Borojerdi, J., Warfel, N. A., Gardner, E. R., Danish, M., Hollander, M. C., Kawabata, S., Tsokos, M., Figga, W. D., Steeg, P. S., and Dennis, P. A. (2007) Nelfinavir, a lead HIV protease inhibitor, is a broad-spectrum, anticancer agent that induces endoplasmic reticulum stress, autophagy, and apoptosis in vitro and in vivo, *Clin. Cancer. Res.* 13, 5183-5194.
 133. Peterson, L. B. (2012) Investigation of the Hsp90 C-terminal binding site, Novel inhibitors and isoform-dependent client proteins, https://kuscholarworks.ku.edu/dspace/bitstream/1808/10218/1/Peterson_ku_0099D_12224_DATA_1.pdf.

CHAPTER 3

Introduction to computational chemistry and molecular modeling

This chapter provides a general introduction to computational chemistry, different molecular modeling, simulation techniques and their applications. Some theoretical descriptions of the computational methods have been explained where it is appropriate. This is accompanied by brief explanations about the various computational tools employed in HIV research with primary focus on molecular dynamics simulations, quantum mechanics, molecular mechanics, molecular docking, binding free energy calculations and QSAR statistical model.

3.1. Introduction to Computational Chemistry

Computational chemistry is a subdivision of chemistry that uses principles of computer skills to aid working out chemical or biochemical problems to reduce the expense and time expended on the discovery, designing and development of new drugs (1). Nevertheless, computer aided drug design (CADD) and computer-aided molecular design (CAMD) can help increase the effectiveness of the drug design process by using the available experimental data (2, 3), thereby detecting lead compounds and predicting their side-effects.

In the research for this thesis, molecular dynamics (MD), one of the two basic methods of computational chemistry was used for the conformations of macromolecules when smaller molecules bind to it. Therefore, this chapter begins with the description of the computational chemistry theories found in both molecular mechanics (MM) and MD: Potential energy surface (PES); Born and Oppenheimer; the Schrödinger's equation, followed by the application of MM, MD and QSAR methods used in our surveys.

3.2. Schrödinger's equation

Quantum mechanics originated from the basic rule employed in daily life when applied to the transfer of very small energies to objects with very low masses. However, in quantum mechanics, all properties of a system are expressed in terms of a wave function that is obtained by solving the Schrödinger's equation. In mathematical physics, the Schrödinger's equation playing the same role as the Hamilton's laws of motion is one of the basic equations in non-relativistic quantum mechanics and non-relativistic classical mechanics respectively. There are

two types of Schrödinger's equation: the time-dependent Schrödinger's equation and the time-independent Schrödinger's equation (4, 5). The widely used Schrödinger's equation in computational chemistry is the time-dependent equation, which defines the Hamiltonian operator as the sum of the kinetic energy and the potential energy.

The general form of the Schrödinger equation is as follows:

$$H = T + V \quad \text{Eq. 1}$$

Where H is the Hamiltonian operator, the sum of the kinetic energy

T is the potential energy

V is the operator

H can also be defined as the following:

$$H = \left[-\frac{\hbar^2}{8\pi^2} \sum_i \frac{1}{m_j} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] + \sum_i \sum_{<j} \left(\frac{e_i e_j}{r_{ij}} \right) \quad \text{Eq. 2}$$

3.3. Born-Oppenheimer approximation

Born and Oppenheimer showed in 1927, after the publication of the Schrödinger equation, how the nuclear motion in a molecule can be determined as a function of its electronic motion. This definition simplifies the Schrödinger's equation for a molecule, that it may be divided into an electron and a nucleic equation where the total internal energy of a molecule can be calculated by solving the electronic Schrödinger's equation. Furthermore, this approximation demonstrated that a molecule has a shape, thus, the nuclei see the electrons as a cloud of negative charge that cover the surface of the molecule. The energy of a molecule is a function of the electron coordinates, but depends on the parameters of the nuclear coordinates which define the molecular geometry i.e. for each geometry, there is a particular energy (6).

$$T^{elec} = \left[-\frac{\hbar^2}{8\pi^2 m} \sum_i^{electrons} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] \quad \text{Eq. 3}$$

Here, the Schrödinger's equation for fixed nuclei electrons is written as:

$$H^{elec} \varphi^{elec}(\mathbf{r}, \mathbf{R}) = E^{eff}(\mathbf{R}) \varphi^{elec}(\mathbf{r}, \mathbf{R}) \quad \text{Eq. 4}$$

Solving this equation for other fixed positions of concern will produce a Potential Energy Surface (PES).

3.4. Potential Energy Surface

A PES is an effective mathematical or graphical representation between the molecular vibrational motion of a molecule and its geometry (7). The concept of a potential energy surface is basic to the quantum mechanical description of molecular energy states and dynamical processes. A PES is basically a plot of the molecular energy versus molecular geometry which enables the understanding of structural characterization (5) as shown in Figure 3.1.

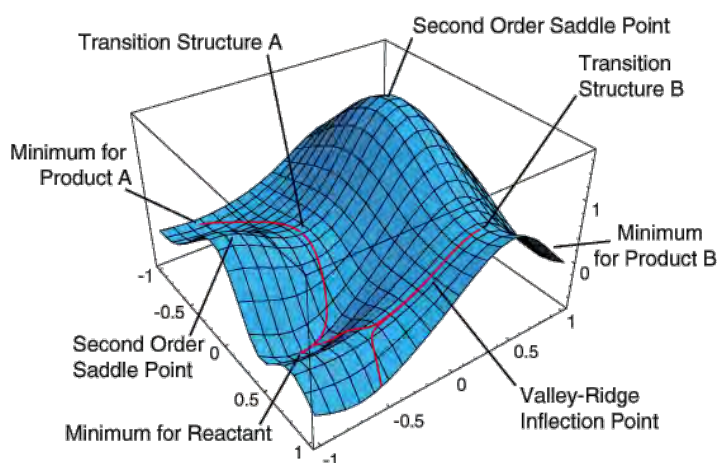


Figure 3.1. A model of a two-dimensional potential energy surface (8).

3.5. Molecular Mechanics

Molecular mechanics (MM) are based on the laws of classical physics to predict the energy of a molecule as a function of its conformation. The conformational analysis, which is probably the most important features of molecular mechanics, enables prediction of the molecule's geometrical and transitional state equilibrium as well as its relative energies between conformers (9). While many of the details of mechanical and biochemical interactions in cellular enzymes are presently uncertain, MM can rely on force-fields with fixed parameters to provide better understanding of the mechanical deformation of DNA, RNA and proteins; changes in cellular structure, response and function (10). This understanding can offer new prognoses of diseases as MM calculations are used to provide qualitative descriptions of

molecular interactions. In spite of the noticeable advantages of molecular mechanics models, it has some setbacks. Firstly, it lacks the ability to provide information about bond formation or bond breaking. Secondly, it is limited to the description of equilibrium geometries and equilibrium conformations. Thirdly, the force fields mostly used nowadays are specific to experimental data being used as parameters; thus, it cannot describe the structures and conformations of new unfamiliar chemical entities that are out of the range of its parameters. Lastly, its inability to predict chemical reactions (11).

The MM total energy of a molecule is defined as the sum of many interactions; bonded interactions such as bond length, bond angles, torsions and non-bonded interactions such as van der Waals and electrostatic (12); as described in the following equation:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} \dots \quad \text{Eq. 5}$$

Where E_{tot} is the total energy, E_{str} is the bond-stretching energy, E_{bend} is the angle-bending energy, E_{tors} is the torsional energy, E_{vdw} is the van der Waals energy and E_{elec} is the electrostatic energy.

This equation together with the parameters required to describe the characteristics of various molecules is termed the force field.

3.5.1. Force fields

A force field is a mathematical function that predicts the energy associated with the conformation of a molecule. It serves as the basic model for molecular mechanics and molecular dynamics calculations. Several molecular mechanics force fields have been developed and are being used today. Force fields such as AMBER, CHARMM, OPLS and GROMOS are the most popular and commonly used set of parameters applied to the simulation of biomolecules. All force fields have different parameters, thus, they must be adjusted to give results of the forces acting within a molecule (12). In our study, AMBER Force Field (GAFF) was used to parameterize the ligands while the standard AMBER force field was used for the human Hsp90 homologue (13). Detailed discussion about the AMBER force field is provided in chapter 5.

3.6. Molecular Dynamics simulations

In a comprehensible way, molecular dynamics is concerned with molecular motion i.e. the characteristic of all chemical processes. Chemical reactions, hormone-receptor binding, and other complex processes are associated with many kinds of intra- and inter-molecular motions. Simulation can be used as a valuable predictive tool for molecular systems and the most widely used simulation methods are Monte Carlo, Brownian dynamics and molecular dynamics. The Monte Carlo method depends on probabilities by generating large numbers of configurations and moving from one state to the other in a specific statistical manner (14). Brownian dynamics is also an effective approach for simulations of large macromolecules (14). However, Molecular dynamics (MD) is the most detailed molecular simulation, in computational chemistry, which is used to calculate the motions of individual molecules (15). It efficiently evaluates different dynamic quantities that cannot be generally obtained by Monte Carlo and it also provide powerful tools for the investigation of the conformational energy of these molecules (13). MD provides a way to test whether the theoretical models predict experimental observations of molecules such as proteins. MD calculations are based on MM principles, and the conformations are created by incorporating newton's laws of motion (16). The result is represented as a trajectory that specifies how the particle position and velocity varies with time (17). Thus, this is achieved by the determination of force (F_i) for each particle as a function of time, which is equal to the negative gradient of the potential energy.

$$F_i = -\frac{U}{r_i} \quad \text{Eq. 6}$$

Where U represents the potential function and r is the position of the particle. According to Newton's Laws of Motion, acceleration (a) of a particle is calculated by dividing the force by the mass of a particle:

$$a_i = -\frac{F_i}{m_i} \quad \text{Eq. 7}$$

The velocity change is equal to the integral of acceleration over time and the change in position is equal to the integral of velocity over time:

$$dv = \int a dt, \quad \text{Eq. 8}$$

$$dr = \int v dt, \quad \text{Eq. 9}$$

Finally, the kinetic energy can be defined in terms of the velocities and momenta of the given particles:

$$K(\mathbf{v}) = \frac{1}{2} \sum_{i=1}^N m_i \mathbf{v}_i \quad \text{Eq. 10}$$

$$K(\mathbf{p}) = \frac{1}{2} \sum_{i=1}^N \frac{p_i^2}{m_i} \quad \text{Eq. 11}$$

Therefore, the total energy of a given system, which is the sum of kinetic and potential energies, is called Hamiltonian (H):

$$H(\mathbf{q}, \mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q}) \quad \text{Eq. 12}$$

Where q is defined as a set of Cartesian coordinates, p is the momenta of the particles and $U(q)$ represents the potential energy function. The velocities, $v_i(t)$, are the first derivatives of the positions with respect to time:

$$\mathbf{v}_i(t) = \frac{d}{dt} \mathbf{q}_i(t) \quad \text{Eq. 13}$$

Here $q_i(t)$ is the atomic positions at a specific time, t . Based on the initial atom coordinates of a particular system, new positions and velocities of the atoms at a given time t and the atoms will migrate to these new positions. Therefore, the new conformations are generated and the temperature of the system is directly proportional to the average kinetic energy (16, 17).

3.7. Approaches for estimating binding affinities

The computer-assisted molecular design (CAMD) approach was employed in this study to optimize new lead compounds for intensive research effort for safe and effective HER2+ breast cancer inhibitors. The advantages of CAMD include: determination of the structure of the host, docking and building a model of the binding site, searching databases for new hosts and the prediction of binding constants or biological activity. The approaches used in this study are explained in the following sub-sections.

3.7.1. Molecular docking

Molecular docking is regularly used in structure-based drug design to identify accurate conformations of ligands to their protein targets and to estimate the strength of the protein-ligand interaction. Molecules such as inhibitors or other drug candidates are identified in the

active site of macromolecules. Macromolecules are proteins such as a receptor, enzyme or nucleic acid with known conformations (18). The ligand-receptor binding energy is calculated as follows:

$$E_{binding} = E_{target} + E_{ligand} + E_{target-ligand} \quad \text{Eq. 14}$$

There are numerous molecular docking programs that are used for academic and commercial purposes (19) such as Dock (20), AutoDock (21), GOLD (22), FlexX (23), GLIDE (24, 25), ICM (19), PhDOCK (26), Surflex (26) and others. These programs can be grouped into four categories: fragment-based, evolutionary-based, stochastic Monte Carlo and the shape complementary methods (27). Although these classes of methods require different information in addition to protein structures, they all share four common computational steps: simplified and rigid body search; selecting the section(s) of interest; modification of docked structures and selecting the best models, respectively (19). Each method is ideal for precise docking problems, however, combining different computational methods can improve the reliability and accuracy of results (28). There are two types of docking: rigid and flexible docking. In rigid docking, the macromolecule and ligand are kept rigid while in flexible docking, flexibility is allowed either for the macromolecule or ligand or both macromolecule and ligand.

The docking method used in this research study is the advanced version of AutoDock, AutoDock Vina (29). This method enables prediction of flexible ligands' binding affinity to rigid human Hsp90 NTD and CTD active sites (29) (detailed discussion in chapter 5). Molecular docking can be applied in different areas including: virtual screening (hit identification), drug discovery (lead optimization), prediction of biological activity, binding-site identification (blind docking), de-orphaning of a receptor, protein-protein interaction, structure-function studies, enzymatic reaction mechanisms and protein engineering. Also, the most typical case which is protein-ligand docking is used to predict the biological activity of a ligand. However, some problems have been reported in Autodock docking (30, 31) which could be due to the following:

- **Posing:** The process of determining whether a given conformation and orientation of a ligand fits the active site.
- **Scoring:** The pose score is a measure of the fit of a ligand into the active site. Scoring during the posing phase usually involves simple energy calculations (electrostatic, van

der Waals, ligand strain). Further re-scoring might attempt to estimate more accurately the free binding energy and perhaps include properties like entropy and solvation.

Implementation of hybrid methods has been reported to create improved algorithms (31). In Chapter 5, we implement the hybrid method by the introduction of MD simulations for determination of ligand stability on the enzyme active site and binding free energy calculations for the validation of docking results.

3.7.2. Binding free energy calculations

The binding free energy calculations enable a comprehensive analysis of the amount of energies that are responsible for molecular stability or binding affinity. The Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) method which is used to estimate binding energies was introduced by Srinivasan *et al.* (32). The MM/PBSA approach is an effective method for the calculation of free energies of molecular systems at reduced computational expenses (33-36). Other studies also documented that MM/GBSA predicts free binding energy better than MM/PBSA (37-39). In the current study, the free binding energy of protease inhibitors (PIs) to the human Hsp90 homologue active site was analyzed using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method (40-43) to validate the docking studies. A brief explanation about free binding energy calculation is provided in chapter 5.

3.7.3. Entropy calculations

Entropy can be defined as the specific information necessary to measure the degree of uncertainty of a thermodynamic system. The setback of all simulation methods is obtaining an accurate entropy value (44). The free energy shows entropic and enthalpy contributions, thus, the values of the entropy and enthalpy show their contribution to the overall free energy of the enzyme-ligand complex (45, 46). The discovery that entropy has an effect on enzyme catalysis has shed more light on the origin of enzyme catalysis (47). A recent study has found an important significant that the activation entropy in enzyme and water are alike and that the overall catalysis is due to enthalpy effects (48).

The simplest way of describing the entropic contribution is that it is the degree of the unpredictability of the overall system (15). According to the Gibb's equation, entropy may be

calculated from ΔG and ΔH . It has been proposed that at a given time, the entropy of a ligand-protein interaction can be expressed as the sum of several contributing factors. Usually, a positive entropy change is an indication that water molecules have been released from the complex. The equation below elaborates the definition of entropy:

$$\Delta G = \Delta H - T\Delta S \quad \text{Eq. 15}$$

Therefore,

$$\Delta S = \frac{\Delta G}{\Delta H - T} \quad \text{Eq. 16}$$

Where ΔG is free binding energy, and ΔH is enthalpy ΔS is entropy and T is the temperature.

3.8. Molecular modeling tools used in this study

3.8.1. Homology modeling

Techniques such as X-ray crystallography, NMR spectroscopy and electron microscopy are used to determine the structure of macromolecules. Among the three major approaches used to predict the 3D structures, homology modeling seems to be the easiest one (49). Homology modeling builds the structure of a protein by using the sequence of a protein whose X-ray crystal structure is known as a building template. A good homology modeling depends on the existence, detection and quality of a known homologue with a known structure. Even though high-resolution structures are best obtained through X-ray crystallography, this approach has encountered some drawbacks over the years. A protein crystal structure requires a lot of experimental time and trial run. Furthermore, some biologically important macromolecules lack X-ray crystal structures or high resolution 3D- structural properties with reference to their protein sequence, thus, to resolve this, homology modeling has been implemented (50). Structures with high-resolution values of 1Å show every atom in highly ordered manner while those with lower resolution value of 3Å or >3Å show only the basic contours of the protein chain. It is therefore advisable to choose a crystal structure whose resolution falls between 1 and 3Å when building a homologue. The research study involving homology modeling can be found in chapter 5 section 3.1. The steps outlined below were followed during the construction of the human Hsp90 homologue [Figure 3.2].

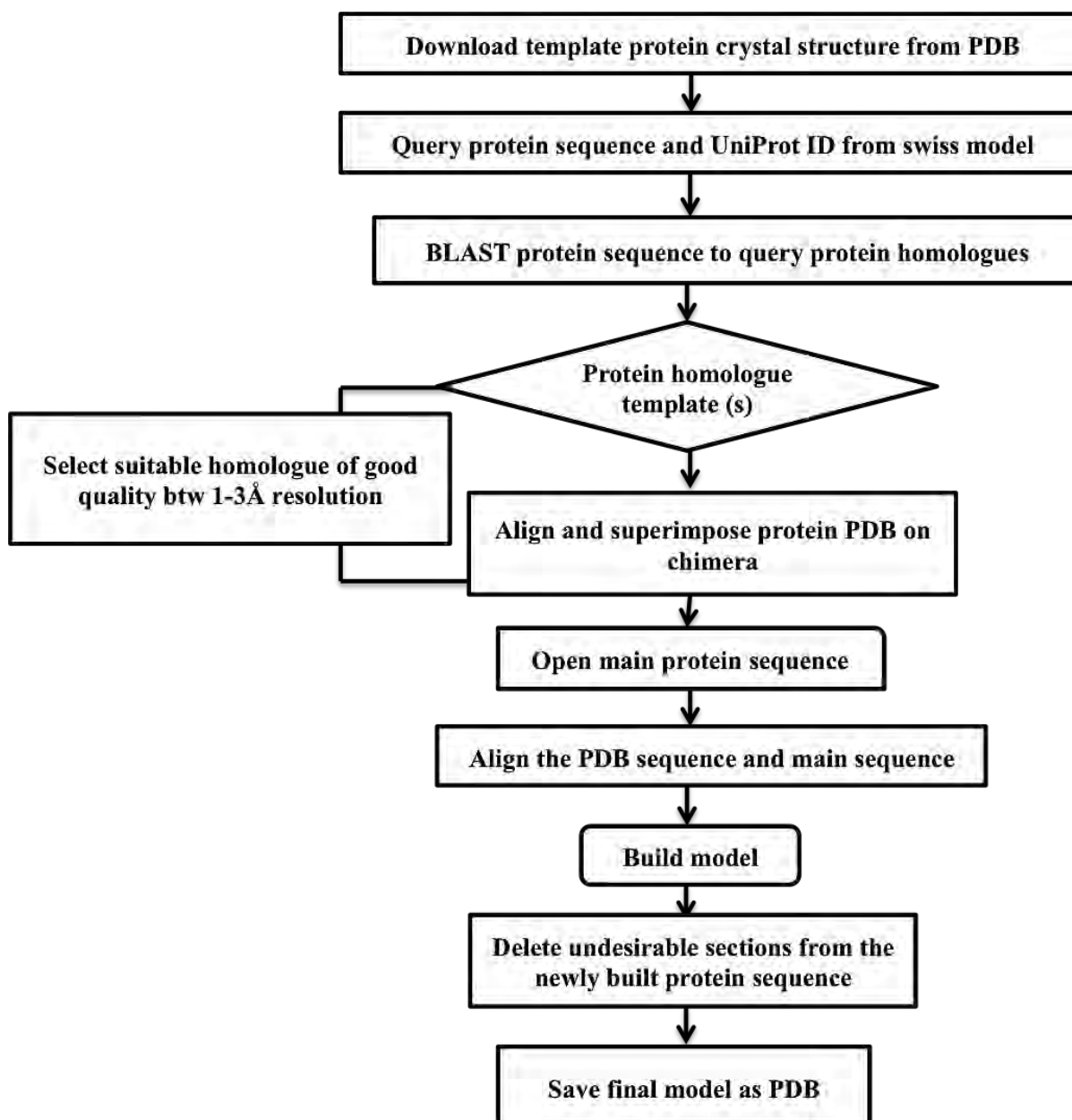


Figure 3.2. Schematic representation of the protocol followed during construction of the human Hsp90 homologue

3.8.2. Per-residue energy decomposition analysis

Per-residue energy decomposition determines the specific energy contributed by each protein residue towards ligand binding. It has also been used to determine bond formation in a molecule-molecule interaction (51-53). It was used to evaluate the residue contributions of the side chain and backbone towards the total binding free energy of a wild type system (54).

3.8.3. Quantitative structural activity relationship (QSAR)

Nowadays, scientists are facing a problem of screening a huge number of molecules in order to predict if they are toxic to human and if they have a significant effect on the target bacteria or virus. In order to avoid the trial-and-error experiments on a huge library of compounds, researchers came up with a process called QSAR to relate molecular characteristics with biological activities whereby a limited number of compounds are synthesized and from their data, derive rules to predict the biological activity of the other compounds. QSAR is a mathematical representation that attempts to correlate the biological activities (IC₅₀ or ED₅₀) and the physicochemical properties of a set of molecules. The QSAR first application was accredited to Hansch *et al.* 1969 (55). The QSAR equation derived by Hansch is given as follows:

$$\log (1/C) = k_1 \log P - k_2(\log P)^2 + k_3 s + k_4 \quad \text{Eq. 17}$$

Where C is the minimum effective dose, logP is the octanol - water partition coefficient, s represents the Hammett substituent constant, k₁₋₃ are the constants derived from regression analysis and k₄ is the constant.

The importance of QSAR include: building a relationship between structure and activity to understand the effect of structure on activity; predictions of biological activities leading to the development of novel compounds; assessment of new chemical entities; determination of stability of a new drug; determination of the distribution of a new chemical entity and lead structure optimization.

A QSAR equation normally takes the general form of a linear equation:

$$\text{Biological Activity} = \text{Const} + (C_1 \times P_1) + (C_2 \times P_2) + (C_3 \times P_3) + \dots \quad \text{Eq. 18}$$

Where the parameters P₁ through P_n are computed for each molecule in the series and the coefficients C₁ through C_n are calculated by fitting variations in the parameters and the biological activity.

$$A = k_1 D_1 + k_2 D_2 + k_3 D_3 + k_n D_n + \text{Const} \quad \text{Eq. 19}$$

Where A is the biological activity, D represents structural properties (descriptors) and K is Regression coefficient.

Some of the descriptors used in building a QSAR model include: Topological (atom counts, ring counts, molecular weights etc.), Geometrical (Molecular volume, surface area etc.), Electronic (Dipole moment, HOMO, LUMO, partial atomic charges etc.) and calculated physical properties (logP, Polarizability, reactive index etc.). The statistical models that are used to derive a QSAR equation include Partial Least Square (PLS), Multiple Linear Regression (MLR), Principal Component Analysis (PCA), Principal Component Regression (PCR) and Genetic Function Algorithm (GFA). The statistical model used in this research study is the GFA method (56) (discussed better in chapter 4). For further reading, some computational studies have recently been reported on HIV-1 integrase inhibitors (57-60) and HER2+ breast cancer inhibitors (61).

References:

1. Ashburn, T. T., and Thor, K. B. (2004) Drug repositioning: Identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov.* 3, 673-683.
2. Talele, T. T., Khedkar, S. A., and Rigby, A. C. (2010) Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic, *Curr. Top. Med. Chem.* 10, 127-141.
3. Song, C. M., Lim, S. J., and Tong, J. C. (2009) Recent advances in computer-aided drug design, *Brief. Bioinform.* 10, 579-591.
4. Rigby, M. (1986) Physical chemistry, 3rd edition - ATKINS, PW, *Nature* 319, 820-820.
5. Lewars, E. G. (2011) Computational Chemistry, *Comp. Theor. Chem.*, 9-43.
6. Tully, J. C. (2000) Perspective on "Zur quantentheorie der molekeln" - Born M, Oppenheimer R (1927) *Ann Phys* 84 : 457, *Theor. Chem. Acc.* 103, 173-176.
7. Wolff, J. J. (1993) Theoretical Aspects of Physical Organic Chemistry. The SN2 Mechanism, *Angew. Chem.* 105, 1277-1278.
8. <http://www.chem.wayne.edu/~hbs/chm6440/PES.html>, 11 Nov. 2013.
9. Lewars, E. G. (2011) *Computational Chemistry*, Springer Netherlands.
10. Bao, G. S., S. (2003) Cell and molecular mechanics of biological materials, *Nature* 2, 715-725.
11. Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems, *Angew. Chem. Int. Ed.* 48, 1198-1229.
12. Guvench, O., and MacKerell, A. D. (2008) Comparison of protein force fields for molecular dynamics simulations, *Methods Mol. Biol.* 443, 63-88.
13. McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins, *Nature* 267, 585-590.
14. Laurent, P., and Titarchuk, L. (1999) Monte-Carlo simulations of comptonization on free-falling electrons, *ApL&C* 38, 173-176.
15. Bronowska, A. K. (2011) Thermodynamics of ligand-protein interactions: Implications for molecular design, In *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*, InTech.
16. Jensen, F. (1999) *Introduction to computational chemistry*, 1st ed., John Wiley & Sons.
17. Lipkowitz, K. B. B., D.B. (1997) *Reviews in computational chemistry*, Vol. 11, VCH Publishers, New York.
18. Vijesh, A. M., Isloor, A. M., Telkar, S., Arulmoli, T., and Hoong-Kun, F. (2013) Molecular docking studies of some new imidazole derivatives for antimicrobial properties, *Arabian J. Chem.* 6, 197-204.
19. Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y., and Humblet, C. (2009) Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy, *J. Chem. Inf. Model.* 49, 1455-1474.
20. Ewing, T. J. A., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases, *J. Comput. Aided Mol. Des.* 15, 411-428.
21. Morris, G. M., Goodsell, D. S., Huey, R., and Olson, A. J. (1996) Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4, *J. Comput. Aided Mol. Des.* 10, 293-304.
22. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD, *Proteins: Struct., Funct., Genet.* 52, 609-623.
23. Warren, G. L., Andrews, C. W., Anna-Maria, C., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D.,

- Woolven, J. M., Peishoff, C. E., and Head, M. S. (2006) A critical assessment of docking programs and scoring functions, *J. Med. Chem.* 49, 5912-5931.
24. 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., and Shenkin, P. S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47, 1739-1749.
25. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem.* 47, 1750-1759.
26. Joseph-McCarthy, D., and Alvarez, J. C. (2003) Automated generation of MCSS-derived pharmacophoric DOCK site points for searching multiconformation databases, *Proteins: Struct., Funct., Genet.* 51, 189-202.
27. Scheraga, H. A., Khalili, M., and Limo, A. (2007) Protein-folding dynamics: Overview of molecular simulation techniques, *ARPC* 58, 57-83.
28. Vajda, S., and Kozakov, D. (2009) Convergence and combination of methods in protein-protein docking, *Curr. Opin. Struct. Biol.* 19, 164-170.
29. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *JCoCh* 19, 1639-1662.
30. Caballero, N. A., Melendez, F. J., Nino, A., and Munoz-Caro, C. (2007) Molecular docking study of the binding of aminopyridines within the K⁺ channel, *J. Mol. Model.* 13, 579-586.
31. Kriegenburg, F., Ellgaard, L., and Hartmann-Petersen, R. (2012) Molecular chaperones in targeting misfolded proteins for ubiquitin-dependent degradation, *FEBS J.* 279, 532-542.
32. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices, *J. Am. Chem. Soc.* 120, 9401-9409.
33. Miller, B. R., III, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An efficient program for end-state free energy calculations, *J. Chem. Theory Comput.* 8, 3314-3321.
34. Homeyer, N., and Gohlke, H. (2012) Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method, *Mol. Inf.* 31, 114-122.
35. Swanson, J. M. J., Henchman, R. H., and McCammon, J. A. (2004) Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy, *BpJ* 86, 67-74.
36. Honig, B., and Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Sci* 268, 1144-1149.
37. Jayaram, B., Sprous, D., and Beveridge, D. L. (1998) Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field, *JPCB* 102, 9571-9576.
38. Bashford, D., and Case, D. A. (2000) Generalized born models of macromolecular solvation effects, *ARPC* 51, 129-152.
39. Ylilauri, M., and Pentikainen, O. T. (2013) MMGBSA As a Tool To Understand the Binding Affinities of Filamin-Peptide Interactions, *J. Chem. Inf. Model.* 53, 2626-2633.
40. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res.* 33, 889-897.

41. Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspect. Drug Discov. Des.* 18, 113-135.
42. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized born solvation model in macromolecular simulations, *Biopolymers* 56, 275-291.
43. Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the generalized Born model suitable for macromolecules, *JPCB* 104, 3712-3720.
44. Ebadi, A., Razzaghi-Asl, N., Khoshneviszadeh, M., and Mini, R. (2013) Comparative amino acid decomposition analysis of potent type I p38 α inhibitors, *DARU*, 1-12.
45. Villà, J., Štrajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., and Warshel, A. (2000) How important are entropic contributions to enzyme catalysis?, *Proc. Natl. Acad. Sci. U. S. A.* 97, 11899-11904.
46. Cao, J. (2011) In silico enzyme modeling: A study of entropic contributions to enzyme catalysis and the design of artificial kemp eliminase, In *Chemistry*, University of Southern California, Los Angeles.
47. Page, M. I. (1977) Entropy, binding-energy, and enzymic catalysis, *Angew. Chem. Int. Ed.* 16, 449-459.
48. Snider, M. J., Gaunitz, S., Ridgway, C., Short, S. A., and Wolfenden, R. (2000) Temperature effects on the catalytic efficiency, rate enhancement, and transition state affinity of cytidine deaminase and the thermodynamic consequences for catalysis of removing a substrate "anchor", *Biochemistry* 39, 9746-9753.
49. Nollmann, M., Stark, W. M., and Byron, O. (2005) A global multi-technique approach to study low-resolution solution structures, *JApCr* 38, 874-887.
50. Pitman, M. R., and Menz, R. I. (2006) Methods for protein homology modelling, *Bioinformatics* 6, 37- 59.
51. Zoete, V., Meuwly, M., and Karplus, M. (2005) Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition, *Proteins: Struct., Funct., Bioinf.* 61, 79-93.
52. Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes, *JMBio* 330, 891-913.
53. Lafont, V., Schaefer, M., Stote, R. H., Altschuh, D., and Dejaegere, A. (2007) Protein-protein recognition and interaction hot spots in an antigen-antibody complex: Free energy decomposition identifies "efficient amino acids", *Proteins: Struct., Funct., Bioinf.* 67, 418-434.
54. Zoete, V., and Michielin, O. (2007) Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: Application to the TCR-p-MHC complex, *Proteins: Struct., Funct., Bioinf.* 67, 1026-1047.
55. Hansch, C., Maloney, P. P., and Fujita, T. (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients, *Nature* 194, 178-180.
56. Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., and Prachayasittikul, V. (2009) A practical overview of quantitative structure-activity relationship, *Excli J.* 8, 74-88.
57. Barreca, M. L., De Luca, L., Ferro, S., Rao, A., Monforte, A.-M., and Chimirri, A. (2006) Computational and synthetic approaches for the discovery of HIV-1 integrase inhibitors, *Arkivoc*, 224-244.

58. Gupta, P., Garg, P., and Roy, N. (2012) Identification of Novel HIV-1 Integrase Inhibitors Using Shape-Based Screening, QSAR, and Docking Approach, *Chem. Biol. Drug Des.* 79, 835-849.
59. Ko Gene, M., Reddy, S. A., Garg, R., Kumar, S., and Hadaegh, A. R. (2012) Computational Modeling Methods for QSAR Studies on HIV-1 Integrase Inhibitors (2005-2010), *Curr. Comput. Aided Drug Des.* 8, 255-270.
60. Sechi, M., Sannia, L., Carta, F., Palomba, M., Dallochio, R., Dessi, A., Derudas, M., Zawahir, Z., and Neamati, N. (2005) Design of novel bioisosteres of beta-diketo acid inhibitors of HIV-1 integrase, *Antiviral Chem. Chemother.* 16, 41-61.
61. Akcay, H. T., and Bayrak, R. (2014) Computational studies on the anastrozole and letrozole, effective chemotherapy drugs against breast cancer, *Spectrochim Acta A Mol Biomol Spectrosc* 122, 142-152.

CHAPTER 4

QSAR study on diketo acid and carboxamide derivatives as potent HIV-1 integrase inhibitor

Arodola Olayide Adebimpe,^a Radha Charan Dash^a and Mahmoud E. S. Soliman^{a*}

^aDiscipline of Pharmaceutical Sciences, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

□ Corresponding authors: Mahmoud E. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 0312607413, Fax: +27 031260 779

Abstract:

Herein, we present a validated predictive QSAR model to provide more insight into the relationship between the molecular properties of diketo acid and carboxamide derivatives as well as HIV-I integrase inhibition. A set of 40 diketo acid and carboxamide derivatives possessing integrase inhibitory activity was subjected to 2D-QSAR using Discovery studio V3.5. The QSAR results presented here were based on a genetic function algorithm (GFA) approach. Logarithmic inverse values of IC_{50} (μM) were taken as the dependent variables, and physicochemical parameters were taken as the independent variable. A suitable set of molecular descriptors was calculated using GFA approach (max 500 generations). Results showed that radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with the correlation coefficient value of 0.820 and play an important role in HIV-1 integrase inhibition.

Key Words: Diketo acid, carboxamide, 2D-QSAR, GFA, Integrase inhibitor

Introduction

Acquired immunodeficiency syndrome (AIDS), currently regarded as one of the most devastating diseases of the human immune system is caused by the human immunodeficiency virus (HIV) (1-3). AIDS was first reported in 1981(4, 5), and now has become a global pandemic. The integration of HIV-1 DNA into the host chromosome contains a series of DNA cutting and joining reactions. The first step in the integration process is 3' end processing. In the second step, termed DNA strand transfer, the previously processed viral DNA end is inserted into the target DNA (6, 7). Thus, the integrase enzyme is crucial for viral replication and represents a potential target for antiretroviral drug design (8-11).

It has been almost forty years since the quantitative structure-activity relationship (QSAR) paradigm first found its way into the practice of pharmaceutical chemistry, toxicology (QSTR) (12), property (QSPR) (13), and eventually most aspects of chemistry. The first application of QSAR is attributed to Hansch *et al.* (14), who developed an equation that related biological activity to certain physicochemical properties of a set of structures (15) and this has become one of the most useful approach to speed up drug design process (16, 17).

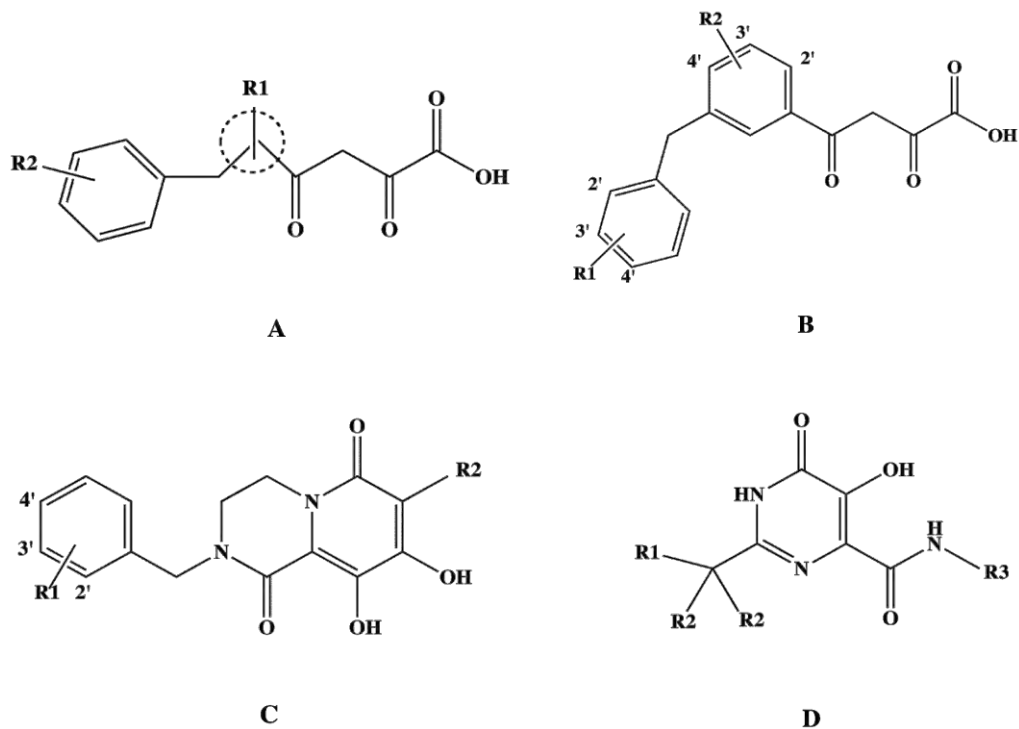
QSAR yield power may be attributed to the strength of its initial postulate that activity was a function of structure as described by electronic attributes and steric properties. QSAR attempts to find consistent relationships between the variations in the values of molecular properties and the biological activity for a series of compounds so that it can be used to evaluate new chemical entities (14). The formulation of thousands of equations using the QSAR methodology attempts a validation of its concepts and its utility in the elucidation of the mechanism of action of drugs at the molecular level and a more complete understanding of physicochemical properties. It is now possible to develop a model for a system as well as compare models from a biological database (18).

There is a series of statistical model analyses that are used to develop a QSAR model, which includes multiple linear regression (MLR), principle component analysis (PCA), partial least square (PLS) and genetic function algorithm (GFA).

In this study, we describe the application of QSAR models based on GFA approach. GFA is a heuristic search method used for identifying optimal solutions to a problem where the possible solution space is too large to be exhaustively enumerated. GFA has been widely used for feature optimization in QSAR models for variable selection (19-21). The purpose of variable selection is to select the variables significantly contributing to prediction and to discard other variables by a fitness function. The GFA approach has a number of important advantages, which include: ability to build multiple models rather than a single model; automatic selection of features to be used in its basic functions and to determine the appropriate number of basic functions to be used by testing full-size models rather than incrementally building them; reliable discovery of combinations of basic functions that take advantage of correlations between features; ability to incorporate the lack of fit (LOF) error measure developed by Friedman (22) that resists over fitting and allows user control over the smoothness of fit; use of larger variety of basic functions in construction of its models, preferred model length and useful partitions of the data set, automatic removal of outliers and finally, provision of additional information not available from other statistical standard regression analysis. The GFA has been applied to three published data sets to demonstrate it is an effective tool for doing both QSAR and QSPR (23-25).

Although several QSAR studies on HIV integrase inhibitors have been reported (26-32) using MLR, PLS and PCA, the QSAR study on HIV-1 integrase using the GFA method has been lacking in literature. Such an understanding about the GFA method might provide a new starting point for the design of novel inhibitors against HIV-1. The main purpose of this work is to find out how accurate the QSAR analysis predicted the activities of compounds that were already synthesized in comparison to their experimental biological activities. Therefore, a 2-dimensional QSAR model was used to analyze some potential diketo acid and carboxamide-based HIV 1 integrase inhibitors.

The list of the structures of 40 inhibitors employed in this study and their experimental inhibitory concentration (IC_{50}) effective against HIV-1 integrase enzyme was taken from literature (33-35) (Table 4.1).

Table 4.1. Structures and biological activity of training and test set studied in this report

Cpd	Core	R1	R2	R3	IC ₅₀ (μ M)	*pIC ₅₀ (μ M)
1	A	Pyrrole	4'-F	-	0.17	0.770
2	A	O-xylene	-	-	5.67	-0.754
3	A	1,2-(CH ₃)-1 <i>H</i> -pyrrole	-	-	0.22	0.658
4 ^a	A	2,3-(CH ₃) thiophene	-	-	0.18	0.745
5	A	2,4-(CH ₃) thiophene	-	-	0.16	0.796
6	A	1,3-(CH ₃)-1 <i>H</i> - pyrrole	-	-	0.5	0.301
7	A	2,5-(CH ₃) thiophene	-	-	0.5	0.301
8 ^a	B	4'-Cl	-	-	1.0	0.000
9	B	3'-F	-	-	0.25	0.602
10	B	-	4'-OCH ₃	-	0.15	0.824
11	B	-	3'-OCH ₃	-	0.14	0.854
12 ^a	C	4'-F	-	-	0.10	1.000
13	C	H	-	-	0.23	0.638
14	C	2'-Cl	-	-	0.37	0.432
15	C	3'-Cl	-	-	0.04	1.398
16 ^a	C	4'-Cl	-	-	0.38	0.420
17	C	4'-F, 3'-Cl	-	-	0.04	1.398

18	C	4'-F	CN	-	0.02	1.699
19	C	4'-F	Br	-	0.03	1.523
20 ^a	C	4'-F	I	-	0.02	1.699
21	D	N(CH ₃) ₃	tetrahydro- 2 <i>H</i> -pyran	4- fluorotoluene	0.002	2.699
22	D	NHCOCH ₃	CH ₃	4- fluorotoluene	0.007	2.155
23	D	NH-SO ₂ -CH ₃	CH ₃	4- fluorotoluene	0.008	2.097
24 ^a	D	NHCO-N(CH ₃) ₂	CH ₃	4- fluorotoluene	0.018	1.745
25	D	NHSO ₂ -N(CH ₃) ₂	CH ₃	4- fluorotoluene	0.012	1.921
26	D	NHCOCO-N(CH ₃) ₂	CH ₃	4- fluorotoluene	0.01	2.000
27	D	NHCOCO-OCH ₃	CH ₃	4- fluorotoluene	0.015	1.824
28 ^a	D	NHCOCO-OH	CH ₃	4- fluorotoluene	0.004	2.398
29	D	N(CH ₃)COCO-N(CH ₃) ₂	CH ₃	4- fluorotoluene	0.015	1.824
30	D	NHCOCO-1,4-(morpholine	CH ₃	4- fluorotoluene	0.02	1.699
31	D	NHCOCO-1,4-(piperazine	CH ₃	4- fluorotoluene	0.026	1.585
32 ^a	D	NHCOCO-N(CH ₃) ₂	CH ₃	1'-ethyl-2',3'- (OCH ₃)	0.021	1.678
33	D	NHCOCO-N(CH ₃) ₂	CH ₃	1'-ethyl-3'-Cl- 4'-F benzene	0.009	2.046
34	D	NHCO-pyridine	CH ₃	4- fluorotoluene	0.02	1.699
35	D	NHCO-pyridazine	CH ₃	4- fluorotoluene	0.015	1.824
36 ^a	D	NHCO-pyrimidine	CH ₃	4- fluorotoluene	0.007	2.155
37	D	NHCO-oxazole	CH ₃	4- fluorotoluene	0.007	2.155
38	D	NHCO-thiazole	CH ₃	4- fluorotoluene	0.008	2.097
39	D	NHCO-1 <i>H</i> Imidazole	CH ₃	4-	0.006	2.222

40 ^a	D	NHCO-1,3,4- oxadiazole	CH ₃	fluorotoluene	0.015	1.824
				4- fluorotoluene		

$$IC_{50} = \text{Biological activity } (\mu\text{M}) \quad *pIC_{50} [\text{Experimental } pIC_{50} \text{ value}] = -\log IC_{50}$$

^a = Compounds were used in test set

Methods

The QSAR study that we conducted was performed on 40 molecules of the diketo acid and carboxamide derivatives (Table 4.1) which had strand transfer data (IC_{50} - molar concentration of the drug leading to 50% inhibition of enzyme Integrase) that was collected from literature (33-35). Out of 40 molecules, 30 were used as a training set and 10 molecules as a test set to evaluate the internal degree of predictivity of the QSAR equation. With the help of CHEM 3D, which has been used in a previous study (32), different 2D structures were drawn (see Table 4.1), followed by the conversion to 3D structures of reasonable conformations using Discovery studio v3.5 software, which has been previously used in a publication (36). A large number of descriptors were also calculated (e.g. ALogP, molecular weight, molar refractivity, dipole moment, heat of formation, Radius of gyration, Wiener index, Zagreb index etc.) (The full list of descriptors is provided in the supplementary material file). Total charge and total formal charge, which are atomistic descriptors, were found to be irrelevant, due to the insignificant values of zero, and were therefore discarded from this study. 2D QSAR analysis was carried out using the genetic function algorithm (GFA) analysis (with 500 maximum generations) (32).

Calculation of pIC_{50}

Reported IC_{50} (μM) for strand transfer values were manually converted into $-\log IC_{50}$ (pIC_{50}) using the formula given below. The term ' pIC_{50} ' is a scale for expressing IC_{50} value exponentially, which normalizes the actual activity using negative logarithmic function.

$$pIC_{50} = -\log IC_{50}$$

Calculation of molecular descriptors

The molecular descriptors were calculated for the data set using QSAR properties and utility of Discovery studio v3.5. Topological (surface area) and constitutional descriptors (AlogP, Dipole, Molecular weight, Energy, Radius of gyration, Wiener index and Zagreb index etc.) were computed (Table 4.2).

Table 4.2. List of descriptors used in this study

Abbreviations	Definition
M	Minimized energy. Gives the energy after a fast minimization procedure using clean force field
W	Wiener index. It is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule)(37)
Z	Zagreb index. It is the sum of the squares of vertex valences
R	Radius of gyration. It is the measures of the size of an object, a surface, or an ensemble of points. It is the root mean square distance of the object's parts from either its center of gravity or a given axis
Ms	Molecular_3D_SASA. It calculates the total solvent accessible surface area for each molecule using a 3D method.(37).

Conclusion

In this study, we screened 26 preselected descriptors for 40 compounds using GFA method. GFA was then used to generate three different 2D-QSAR models to determine the degree of predictivity of these diketo acid and carboxamide derivatives as HIV-1 integrase inhibitors. A QSAR model was generated for integrase activity. As all the descriptors were not important for specific model generation, in order to select the optimal set of descriptors, we used systematic variable selection leave one out (LOO) method in a

stepwise forward manner for the selection of descriptors. Three best QSAR equations models generated for this study using the GFA approach and LOO method are shown in Table 4.3.

Table 4.3. The 3 different equations derived from the QSAR model

Equation	R ²	Q ²	LOF	P-value
1 Y= -11.65 – 0.0024929W + 0.088809Z + 0.01936M + 1.1879R	0.820	0.558	0.193	5.174e-09
2 Y= -12.896 – 0.0028585W + 0.077907Z + 0.020068M + 0.015681Ms	0.812	0.470	0.202	9.270e-09
3 Y= -9.6736 – 0.0020098W + 0.078883Z + 0.89779R	0.790	0.620	0.190	5.641e-09

Y: pIC₅₀, set of descriptors (W, Z, M, R, Ms,) have been explained in table 2, R²: correlation coefficient, Q²: cross-validated R squared, LOF: Lack of fit, P-value: significance level (32).

The statistical quality of the generated models was determined by the parameters like correlation co-efficient (R²), cross-validated squared correlation co-efficient (Q²), LOF, which is the relative measure of quality of fit, p-value which represents the variance of calculated and observed activity, and chance statistics assuring that the results are not merely based on chance correlations. Best models were selected on the basis of their statistical significance.

Equation 1: The QSAR result for integrase inhibition produced highly predictive model that has excellent R² = 0.820 (see Figure 4.1). Furthermore, the developed model showed Q² = 0.558 and LOF value 0.193 indicating the model is well validated and explained. The four best descriptors selected on the basis of importance in biological activity were Zagreb index (Z), Wiener index (W), Radius of gyration (R) and minimized energy (M).

From the QSAR study, the statistically significant equation derived was,

$$\text{pIC}_{50} = -11.65 - 0.0024W + 0.089Z + 0.019M + 1.187R$$

N= number of training compounds =30, R² = 0.820; Q² = 0.558; LOF = 0.193.

From the equation it was observed that, the radius of gyration and minimized energy and Zagreb index are positively correlated with the biological activity, however radius of gyration had the most contribution towards the biological activity. The radius of gyration of a molecule describes its dimensions and is calculated as the root mean square distance between its centre of gravity and its ends. The value of R in compounds **30, 34, 35** and **36** having amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary the value of R in the compounds **13, 17** and **19** having a halogenated group is low and resulted into lower biological activity. The second parameter Z contributed relatively lower in the above QSAR study but positively correlated with biological activity. The Z is a topological descriptor based on the vertex degree of heavy atoms. It also interprets the activity of compounds **18** and **20** having heavy atoms compared to compounds **9** and **13**. The W is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule. The value of W in the compounds **1, 3,** and **5** having a heavy atom attached to one of the aromatic rings is less. This resulted in lower biological activity. On the other hand, the high value of W in the compounds **30, 34** and **36** is due to the fact that these compounds have one or more amide group attached to one or more cyclic aromatic ring, which gives room for an addition or substitution reaction and thus an increased biological activity.

Equation 2: The QSAR result for integrase inhibition produced a good predictive model that also has R² = 0.812 (see Figure 4.2). Furthermore the developed model showed Q² = 0.470 and LOF value 0.202 indicating the model is also well validated and explained. The four best descriptors selected on the basis of importance in biological activity were Wiener index (W), Zagreb index (Z), minimized energy (M) and total solvent accessible surface area (Ms). From the QSAR study, the statistically significant equation derived was:

$$\text{pIC}_{50} = -12.896 - 0.0028585W + 0.077907Z + 0.020068M + 0.015681Ms$$

N= number of training compounds =30, R² = 0.812; Q² = 0.470; LOF = 0.202.

From the equation it was observed that, the total solvent accessible surface area and minimized energy and Zagreb index positively correlated negatively with biological activity, however Zagreb index contributed mostly towards the biological activity. The

Zagreb index of a molecule describes the sum of squares of vertex valences. The value of Z in compounds **21**, **30**, **31** and **40** having amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary, the value of Z in the compounds **4**, **5** and **7** having a sulfide group attached to an aromatic cyclic ring is less and resulted into lower biological activity. The second parameter M contributed relatively low in the above QSAR but positively correlated with biological activity. The M is a topological descriptor that best describe the energy of a system after a fast minimization procedure and it interprets the high activity of compounds **26** and **29** respectively.

Equation 3: The QSAR result for integrase inhibition produced a good predictive model that also has good $R^2 = 0.790$ (see Figure 4.3). Furthermore the developed model showed $Q^2 = 0.620$ and LOF value 0.190 indicating the model is also well validated and explained. The three best descriptors selected on the basis of importance in biological activity were Wiener index (W), Zagreb index (Z), Radius of gyration (R). From the QSAR study, the statistically significant equation derived was:

$$pIC_{50} = -9.6736 - 0.0020098W + 0.078883Z + 0.89779R$$

N= number of training compounds =30, $R^2 = 0.790$; $Q^2 = 0.620$; LOF = 0.190.

From the equation it was observed that the Zagreb index (Z) and Radius of gyration (R) positively correlated with biological activity, however R contributed mostly towards the biological activity. The value of R in compounds **30**, **34**, **35** and **36** having at least an amide residue attached to a cyclic aromatic ring is high. These compounds showed good activity. On the contrary, the value of R in compounds **13**, **17** and **19** having halogenated group is lower and this resulted in decreased biological activity. The second descriptor Z gave the least contribution in the above equation but positively correlated with the biological activity. Z also interprets the activity of compounds **18** and **20** having heavy atoms compared to compounds **9** and **13**, which shows lower biological activity.



Figure 4.1. Predicted pIC_{50} against experimental pIC_{50} for equation 1

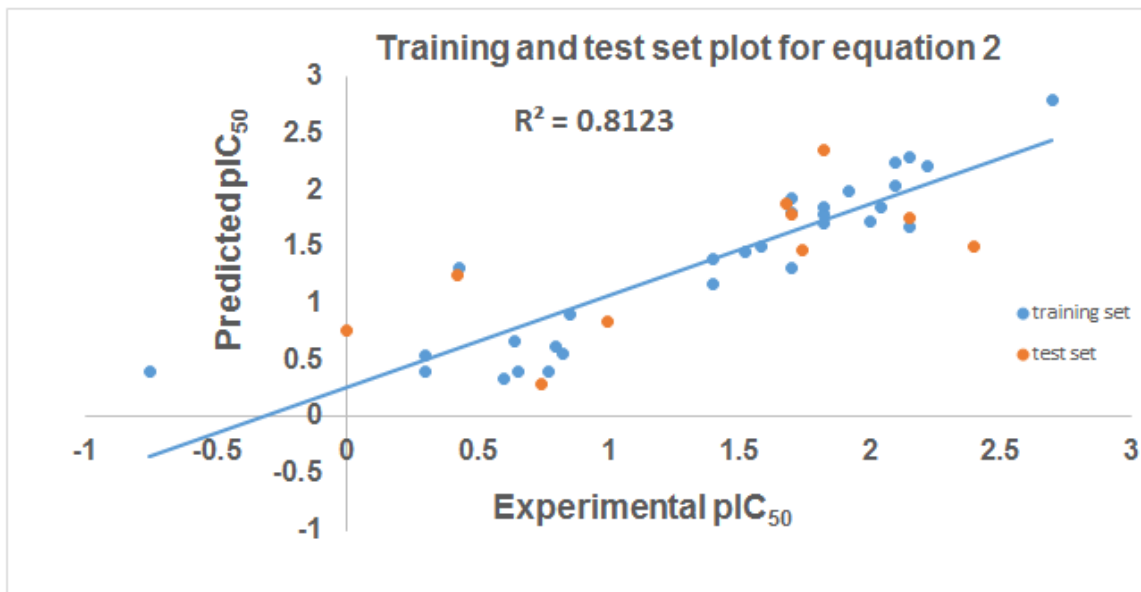


Figure 4.2. Predicted pIC_{50} against experimental pIC_{50} for equation 2



Figure 4.3. Predicted pIC_{50} against experimental pIC_{50} for equation 3

Table 4.4. Experimental pIC_{50} and GFA predicted pIC_{50} for training set

Cmpds	pIC_{50}	Predicted ¹	Residual ¹	Predicted ²	Residual ²	Predicted ³	Residual ³
1	0.77	0.409	0.361	0.393	0.377	0.274	0.496
2	-0.754	0.105	-0.859	0.407	-1.161	0.335	-1.089
3	0.658	0.377	0.281	0.397	0.261	0.261	0.397
5	0.796	0.498	0.298	0.618	0.178	0.228	0.568
6	0.301	0.616	-0.315	0.536	-0.235	0.422	-0.121
7	0.301	0.608	-0.307	0.398	-0.097	0.512	-0.211
9	0.602	0.463	0.139	0.330	0.272	0.602	0.000
10	0.824	0.505	0.319	0.563	0.261	0.692	0.132
11	0.854	0.591	0.263	0.900	-0.046	0.725	0.129
13	0.638	0.971	-0.333	0.676	-0.038	1.017	-0.379
14	0.432	1.280	-0.848	1.316	-0.884	1.276	-0.844
15	1.398	1.239	0.159	1.166	0.232	1.260	0.138
17	1.398	1.267	0.131	1.401	-0.003	1.340	0.058
18	1.699	1.580	0.119	1.311	0.388	1.559	0.139

19	1.523	1.276	0.247	1.464	0.059	1.362	0.160
21	2.699	2.495	0.204	2.796	-0.097	2.334	0.365
22	2.155	1.681	0.474	1.672	0.483	1.713	0.442
23	2.097	1.973	0.124	2.034	0.063	1.989	0.108
25	1.921	1.957	-0.036	1.998	-0.077	1.975	-0.054
26	2.000	1.704	0.296	1.724	0.276	1.777	0.223
27	1.824	1.797	0.027	1.707	0.117	1.867	-0.043
29	1.824	1.943	-0.119	1.851	-0.027	1.883	-0.059
30	1.699	1.970	-0.271	1.926	-0.227	1.929	-0.230
31	1.585	1.391	0.194	1.499	0.086	1.594	-0.009
33	2.046	1.739	0.307	1.845	0.201	1.860	0.186
34	1.699	2.020	-0.321	1.809	-0.110	2.154	-0.455
35	1.824	1.931	-0.107	1.787	0.037	2.017	-0.193
37	2.155	2.325	-0.170	2.302	-0.147	2.090	0.065
38	2.097	2.221	-0.124	2.243	-0.146	2.109	-0.012
39	2.222	2.357	-0.135	2.219	0.002	2.133	0.089

Table 4.4 shows the experimental pIC_{50} and the predicted pIC_{50} using the GFA approach for the training set. This shows how the GFA method predicted the pIC_{50} .

Table 4.5. Experimental pIC_{50} and GFA predicted pIC_{50} for test set

Cmpds	pIC_{50}	Predicted ¹	Residual ¹	Predicted ²	Residual ²	Predicted ³	Residual ³
4	0.745	0.326	0.419	0.287	0.458	0.282	0.463
8	0.000	0.485	-0.485	0.761	-0.761	0.587	-0.587
12	1.000	1.178	-0.178	0.836	0.164	1.215	-0.215
16	0.420	1.212	-0.792	1.259	-0.839	1.233	-0.813
20	1.699	1.482	0.217	1.784	-0.085	1.473	0.226
24	1.745	1.580	0.165	1.471	0.274	1.634	0.111
28	2.398	1.594	0.804	1.500	0.898	1.706	0.692
32	1.678	1.937	-0.260	1.877	-0.199	1.961	-0.283
36	2.155	1.936	0.219	1.765	0.390	2.096	0.059
40	1.824	2.656	-0.832	2.360	-0.536	2.371	-0.547

Predicted¹: predicted value for equation 1, Residual¹: residual value for equation 1

Predicted²: predicted value for equation 2, Residual²: residual value for equation 2

Predicted³: predicted value for equation 3, Residual³: residual value for equation 3

Table 4.5 also shows how well the pIC_{50} was predicted using the GFA approach for the test set. From the residual values (Figures 4.4, 4.5 and 4.6), it can be clearly seen that the lower residual values show that there is a minimal difference between the experimental value and the predicted value of the biological activity of this test set.

The histograms of the three different QSAR models are represented below.

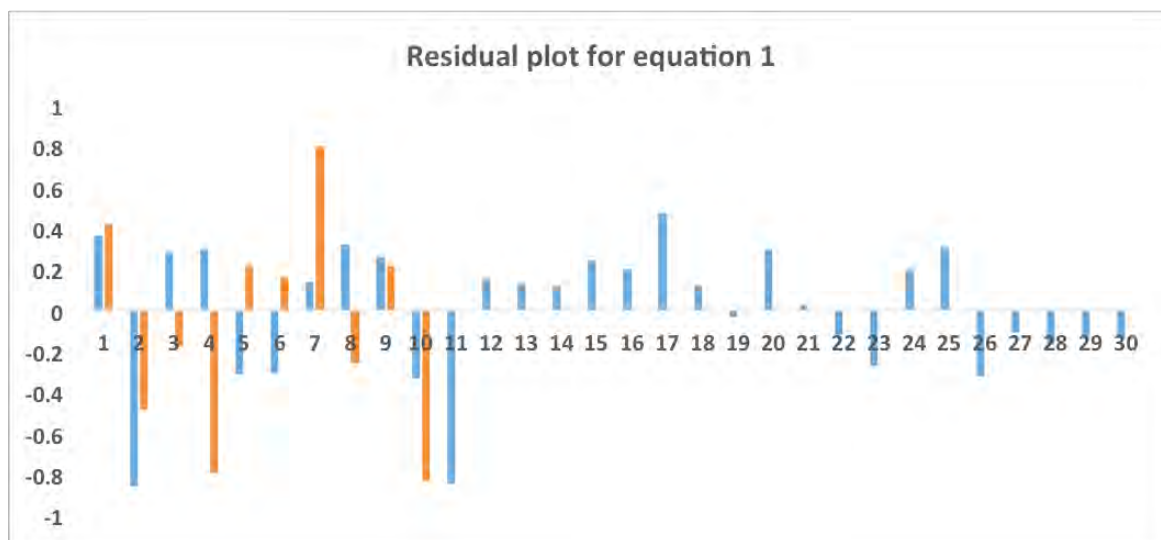


Figure 4.4. Histogram of residual values obtained from QSAR model for equation 1

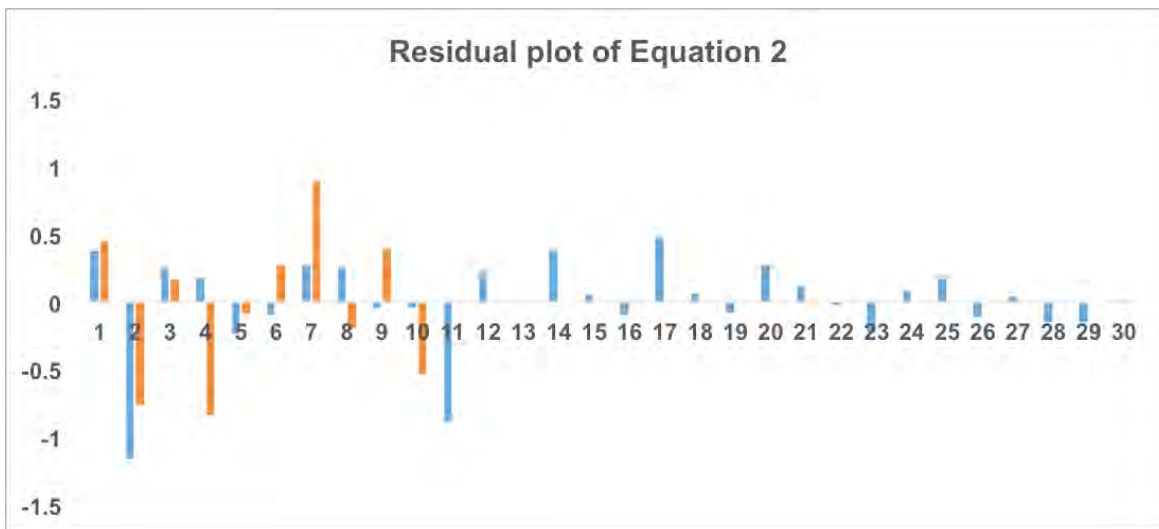


Figure 4.5. Histogram of residual values obtained from QSAR model for equation 2

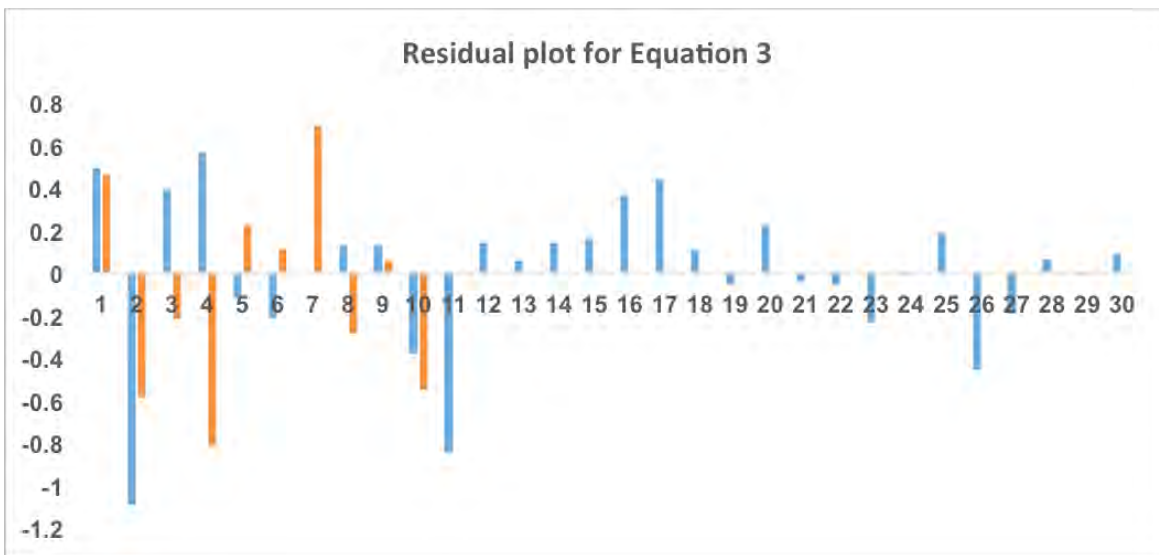


Figure 4.6. Histogram plot of residual values obtained from QSAR model for equation 3

In Figure 4.6, the Y-axis represents the different molecular descriptors used in this study as shown on the right side of the graph. On the other hand, the X-axis represents the number of generations we could generate for each of these molecular descriptors.

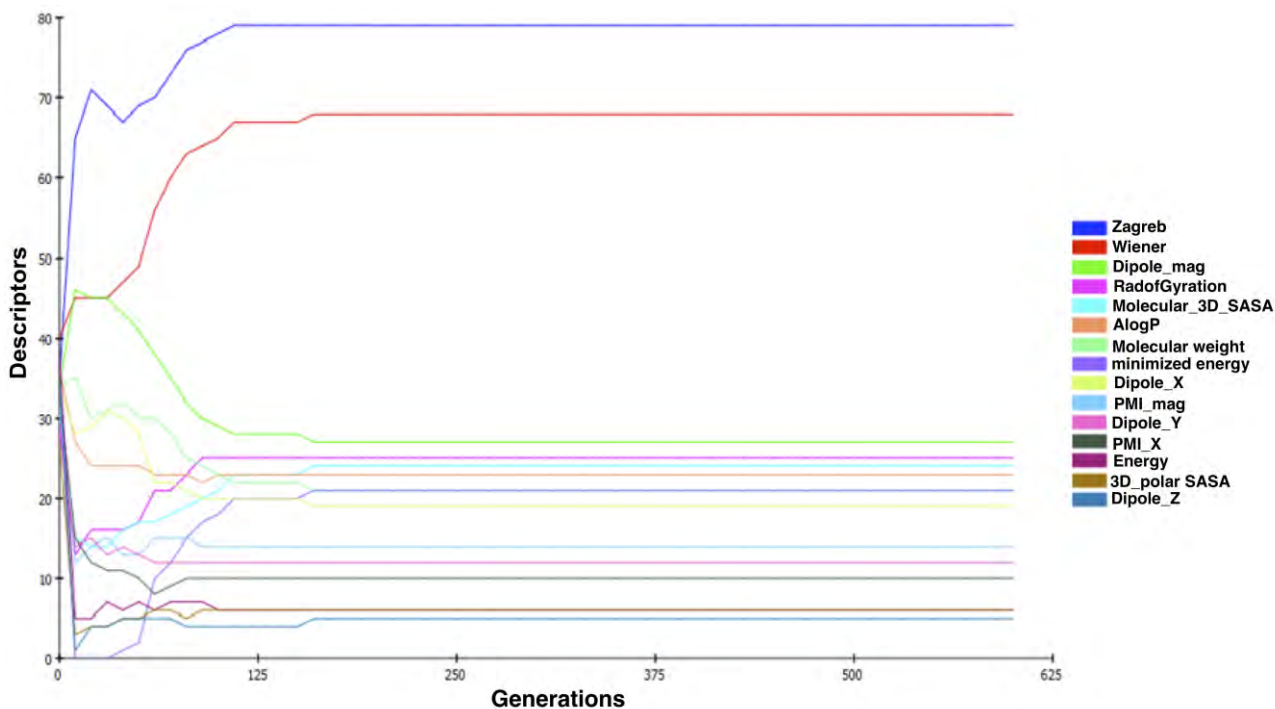


Figure 4.7. The graph of the variable usage against generation number

According to Figure 4.7, at each step, the GFA uses the current population to create the children that make up the next generation. The algorithm selects a group of individuals in the current population, called parents, who contribute their genes—the entries of their vectors—to their children. The algorithm usually selects individuals that have better fitness values as parents. User can specify the function that the algorithm uses to select the parents. The GFA creates three types of children for the next generation: Elite children, Crossover children, and Mutation children. In our QSAR study, the algorithm stops when the number of generations reaches the value of 500 Generations.

In this present study, QSAR models have been developed based on molecular, structural, physicochemical, 2D and 3D properties that were obtained from various softwares. The

Discovery studio result suggests that the Radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with a correlation coefficient value of 0.8209, which is highly significant. These descriptors have played an important role in identifying some promising compounds that possess HIV-1 inhibitory properties such as compound **18**, **20**, **30**, **34**, **35** and **36**. The synthesis of the compounds considered in this study was done in literature (33-35) but it was validated using 2D-QSAR model. This model holds good predictive performance with Q^2 values ranging from 0.47 to 0.62 that was calculated using LOO method. In conclusion, this model can be used to predict more potent drugs that possess HIV-1 integrase inhibitory property.

Acknowledgement

The author acknowledges the computational facility provided by CHPC, Cape Town (<http://www.chpc.ac.za>) and the School of Health Sciences, UKZN, for financial support.

Conflict of Interest

Authors declare no conflict of interest

References:

1. Mitsuya, H., Guo, H., Megson, M., Trainer, C., Reitz jr, M., and Broder, S. (1984) Transformation and cytopathogenic effect in an immune human t-cell clone infected by HTLV-I, *Sci* 223, 1293 - 1296.
2. Kent A. Sepkowitz, M. D. (2001) AIDS - The first 20 years, *N. Engl. J. Med.* 344, 1764 - 1772.
3. Weiss, R. A. (1993) How does HIV cause AIDS?, *Sci* 260, 1273 - 1279.
4. Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981) Pneumocystis-carinii pneumonia and mucosal candidiasis in previously healthy homosexual men - evidence of a new acquired cellular immunodeficiency, *N. Engl. J. Med.* 305, 1425-1431.
5. Masur, H., Michelis, M. A., Greene, J. B., Onorato, I., Vandestouwe, R. A., Holzman, R. S., Wormser, G., Brettman, L., Lange, M., Murray, H. W., and Cunninghamrundles, S. (1981) An outbreak of community-acquired pneumocystis-carinii pneumonia - initial manifestation of cellular immune dysfunction, *N. Engl. J. Med.* 305, 1431-1438.
6. Craigie, R. (2001) HIV integrase, a brief overview from chemistry to therapeutics, *J. Biol. Chem.* 276, 23213-23216.
7. Di Santo, R., Costi, R., Artico, M., Tramontano, E., La Colla, P., and Pani, A. (2003) HIV-1 integrase inhibitors that block HIV-1 replication in infected cells. Planning synthetic derivatives from natural products, *Pure Appl. Chem* 75, 195-206.
8. Barreca, M. L., Ferro, S., Rao, A., De Luca, L., Zappala, M., Monforte, A. M., Debyser, Z., Witvrouw, M., and Chimirri, A. (2005) Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors, *J. Med. Chem.* 48, 7084-7088.
9. d'Angelo, J., Mouscadet, J. F., Desmaële, D., Zouhiri, F., and Leh, H. (2001) HIV-1 integrase: the next target for AIDS therapy?, *Pathol. Biol.* 49, 237-246.
10. Grobler, J. A., Stillmock, K., Hu, B. H., Witmer, M., Felock, P., Espeseth, A. S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J. S., Young, S., Vacca, J., and Hazuda, D. J. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes, *Proc. Natl. Acad. Sci. U.S.A.* 99, 6661-6666.
11. Zhuang, L. C., Wai, J. S., Embrey, M. W., Fisher, T. E., Egbertson, M. S., Payne, L. S., Guare, J. P., Vacca, J. P., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Witmer, M. V., Moyer, G., Schleif, W. A., Gabryelski, L. J., Leonard, Y. M., Lynch, J. J., Michelson, S. R., and Young, S. D. (2003) Design and synthesis of 8-hydroxy- 1,6 naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells, *J. Med. Chem.* 46, 453-456.
12. Aboul-Kassim, Tareka, T., and Simoneit, B., R. T. . (2001) QSAR/QSPR and multicomponent joint toxic effect modeling of organic pollutants at aqueous-solid phase interfaces, *Handb. Environ. Chem.* 5E, 243-314.
13. Das, R. N., and Roy, K. (2013) Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future, *Molec. Divers.* 17, 151-196.

14. Hansch, C., Maloney, P. P., and Fujita, T. (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients, *Nature* *194*, 178.
15. Yang, G.-F., and Huang, X. (2006) Development of quantitative structure-activity relationships and its application in rational drug design, *Curr. Pharm. Des.* *12*, 4601-4611.
16. Hemmateenejad, B., Miri, R., Akhond, M., and Shamsipur, M. (2002) Quantitative structure-activity relationship study of recently synthesized 1,4-dihydropyridine calcium channel antagonists. Application of the Hansch analysis method, *Arch. Pharm. (Weinheim)*. *335*, 472-480.
17. Lill, M. A. (2007) Multi-dimensional QSAR in drug discovery, *Drug Discov Today* *12*, 1013-1017.
18. Hansch, C., Kurup, A., Garg, R., and Gao, H. (2001) Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms, *Chem. Rev.* *101*, 619-672.
19. Cho, S. J., and Hermsmeier, M. A. (2002) Genetic algorithm guided selection: Variable selection and subset selection, *J. Chem. Inf. Comput. Sci.* *42*, 927-936.
20. Fernandez, M., Caballero, J., Fernandez, L., and Sarai, A. (2011) Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM), *Molec. Divers.* *15*, 269-289.
21. Hasegawa, K., Miyashita, Y., and Funatsu, K. (1997) GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists, *J. Chem. Inf. Comput. Sci.* *37*, 306-310.
22. Friedman, J. H. (1991) Multivariate adaptive regression splines, *Ann. Stat.* *19*, 1-67.
23. Selwood, D. L., Livingstone, D. J., Comley, J. C. W., Odowd, A. B., Hudson, A. T., Jackson, P., Jandu, K. S., Rose, V. S., and Stables, J. N. (1990) Structure-activity-relationships of antifilarial antimycin analogs - a multivariate pattern-recognition study, *J. Med. Chem.* *33*, 136-142.
24. Cardozo, M. G., Iimura, Y., Sugimoto, H., Yamanishi, Y., and Hopfinger, A. J. (1992) QSAR analyses of the substituted indanone and benzylpiperidine rings of a series of indanone benzylpiperidine inhibitors of acetylcholinesterase, *J. Med. Chem.* *35*, 584-589.
25. Koehler, M. G., and Hopfinger, A. J. (1989) Molecular modelling of polymers: 5. Inclusion of intermolecular energetics in estimating glass and crystal-melt transition temperatures, *Polymer* *30*, 116-126.
26. Almerico, A. M., Tutone, M., Ippolito, M., and Lauria, A. (2007) Molecular modelling and QSAR in the discovery of HIV-1 integrase inhibitors, *Curr Comput Aided Drug Des* *3*, 214-233.
27. de Melo, E. B., and Castro Ferreira, M. M. (2009) Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* *44*, 3577-3583.
28. Dessalew, N. (2009) Investigation of the structural requirement for inhibiting HIV integrase: QSAR study, *Acta Pharm.* *59*, 31-43.

29. Gupta, P., Roy, N., and Garg, P. (2009) Docking-based 3D-QSAR study of HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* *44*, 4276-4287.
30. Leonard, J. T., and Roy, K. (2008) Exploring molecular shape analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors, *Eur. J. Med. Chem.* *43*, 81-92.
31. Sharma, H., Cheng, X., and Buolamwini, J. K. (2012) Homology Model-Guided 3D-QSAR Studies of HIV-1 Integrase Inhibitors, *J. Chem. Inf. Model.* *52*, 515-544.
32. Srivastav, V. K., and Tiwari, M. (2013) QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors, *Arabian J. Chem. In Press*, <http://www.sciencedirect.com/science/article/pii/S1878535213000270>
33. Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Paz, O. G., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., and Rowley, M. (2008) Discovery of Raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, *J. Med. Chem.* *51*, 5843-5855.
34. Wai, J. S., Egbertson, M. S., Payne, L. S., Fisher, T. E., Embrey, M. W., Tran, L. O., Melamed, J. Y., Langford, H. M., Guare, J. P., Zhuang, L. G., Grey, V. E., Vacca, J. P., Holloway, M. K., Naylor-Olsen, A. M., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Schleif, W. A., Gabryelski, L. J., and Young, S. D. (2000) 4-aryl-2,4-dioxobutanoic acid inhibitors of HIV-1 integrase and viral replication in cells, *J. Med. Chem.* *43*, 4923-4926.
35. Wai, J. S., Kim, B., Fisher, T. E., Zhuang, L., Embrey, M. W., Williams, P. D., Staas, D. D., Culberson, C., Lyle, T. A., Vacca, J. P., Hazuda, D. J., Felock, P. J., Schleif, W. A., Gabryelski, L. J., Jin, L., Chen, I. W., Ellis, J. D., Mallai, R., and Young, S. D. (2007) Dihydroxypyridopyrazine-1,6-dione HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett.* *17*, 5595-5599.
36. Ouyang, L., He, G., Huang, W., Song, X., Wu, F., and Xiang, M. (2012) Combined Structure-Based Pharmacophore and 3D-QSAR Studies on Phenylalanine Series Compounds as TPH1 Inhibitors, *Int. J. Mol. Sci.* *13*, 5348-5363.
37. Karelson, M., Lobanov, V. S., and Katritzky, A. R. (1996) Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.* *96*, 1027-1043.

CHAPTER 5

Could the FDA-approved anti-HIV PR inhibitors be promising anticancer agents? An answer from molecular dynamics analyses

Mbatha H. Sbongile^a, Olayide A. Arodola^a, and Mahmoud E. S. Soliman^{a*}

^aSchool of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

* Corresponding author: Mahmoud E.S. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 031 260 7413, Fax: +27 031 260 779

Abstract

Based on experimental data, the anticancer activity of FDA-approved HIV-1 protease inhibitors (PIs) were previously reported. Heat Shock Protein 90 has been identified as a known anticancer therapeutic target. The experimental data, which shows that nelfinavir has an anticancer inhibitory effect on yeast Hsp90, since it exhibited the highest IC₅₀ value, raised the possibility that NFV might have the same effect on a mammalian Hsp90. The lack of the X-ray crystal structure of human Hsp90 makes the mechanism of binding of these drugs to the enzyme more ambiguous - especially with the existence of more than one possible binding domain in the Hsp90 enzyme. To this end, in this work, we embarked on various computational approaches to investigate the binding mode of the current FDA-approved HIV-1 PIs against Hsp90. Since the X-ray crystal structure of the human Hsp90 protein is not yet resolved, homology modeling was performed to create its 3D structure for subsequent simulations. The two possible binding sites, C-terminal and N-terminal domains, were considered in this study. Eighteen 5 ns molecular dynamic simulations and free binding energy calculations were carried out. Based on the thermodynamics calculations, it was found that these inhibitors are most likely to bind at the N-terminal domain – with a significant binding affinity difference (~ 54.7-83.03 kcal/mol) when compared to C-terminal domain. To our knowledge, this is the first account of detailed computational investigations aimed to understand the binding mechanism of HIV PIs binding to Hsp90. Information gained from this study should also provide a route map towards the design and optimisation of potential derivatives of PIs to treat HER2+ breast cancer.

Keywords: Binding free energy, molecular dynamics, HIV-1 protease inhibitors, anticancer

1. Introduction

Cancer, a heterogeneous disease, is one of the major cause of death worldwide (1). In 2008, GLOBOCAN estimated about 12.7 million cancer cases and 7.6 million cancer deaths (about 21,000 cancer deaths a day) occurrence; of these, 56% (2.8 million) of the cases and 64% (4.8 million) of the deaths occurred in the economically developing countries (1). An estimated 14.1 million cancer cases occurred in 2012 (2). In 2014, an estimated 585,720 cancer deaths is expected in the United States (3). According to the American Cancer Society, it has been reported that there remains an expected increase in the cancer incidences in the next twenty years, more especially in African regions (4, 5).

Cancer is an abnormal proliferation of cells that can lead to malignancy and death. These cells have the potential to elude other normal cells through the process called metastasis (5). Causal factors like tobacco, chemicals, radiation, infectious organisms, inherited mutations and hormones may act together to initiate or accelerate carcinogenesis. Among the numerous types of cancer, breast cancer is the most common cancer diagnosed in women worldwide (6). Approximately two-thirds of breast cancer tumors are hormone dependent, requiring estrogens to grow and estrogens are formed in the human body via a multistep route starting from cholesterol (7).

Estrogen receptor (ER) and human epidermal growth factor 2 (HER2) are familial hormone sensitive receptors of the human epidermal growth factor (EGFR). HER2 consists of four subtypes (HER1–4) (8); which are regulated at the level of expression by hormones; thus, HER2 is of greatest significance as it is a pivotal receptor target for breast cancer treatment. Compared to other types of breast cancer, HER2+ breast cancer tends to be more aggressive and less receptive to hormone treatments. According to published data, it was documented that approximately 25–30% of human breast cancers overexpress HER2 (9, 10), in which heat shock protein (HSP) has been reported to play a role in the overexpression of HER2 (11,

12). Therefore, to block HER2 overexpression, the subsequent step would be to inhibit the enzyme responsible for producing this estrogen, Hsp90. Hsp90 functions as a molecular chaperone, regulating the proper functionality of other proteins in the body (13).

Hsp90 is a homodimer consisting of α and β chains. Each monomer of Hsp90 dimer contains four domains: 1) a highly conserved *N*-terminal (NTD), 2) *C*-terminal domain (CTD), 3) a middle domain (MD), and 4) a charged linker that connects the *N*-terminal and middle domain (14) as shown in Figure 5.1.

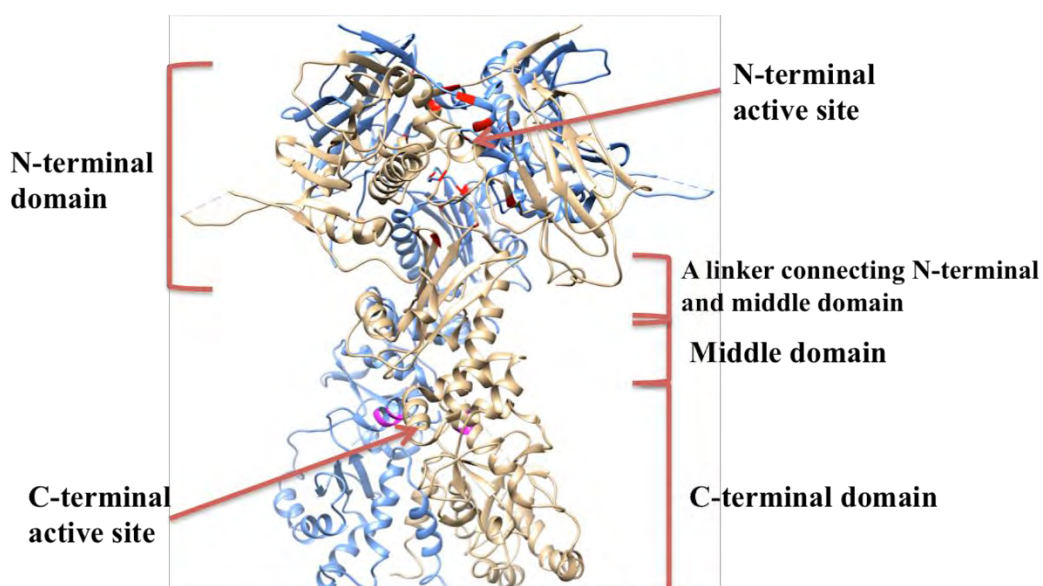


Figure 5.1. The crystal structure of Hsp90 Alpha (blue) and Beta chain (gold) (PDB code: 2CG9) showing its different domains (15).

Hsp90 contains a highly conserved ATP binding site near its NTD, as Hsp90 functions are energetically expensive (16). This Hsp90-ATP binding site has been under intense pharmaceutical investigation, as many designed drugs offer competitive inhibition against ATP for this site (14). The design, development and discovery of new cancer chemotherapeutics cannot meet the demand and rate at which the disease is manifesting. The devastating effects are evident from the high mortality rate of inflicted patients, due to

progression of the disease or inherited infection contracted from having a compromised immunity. Recent studies projects that a more efficient procedure for drug development that reduces costs is to find new indications for already approved drugs; this process is referred to as “repositioning” (17-19). “Repositioning” takes advantage of available data on existing drugs, limits risks and costs to pharmaceutical companies, and could advance the evaluation and movement of new cancer therapies to the clinic.

The study conducted by Shim *et al.*, analyzed the proteolytic profile of yeast Hsp90 α full length for both *N*- and *C*- terminal domains. This study identified five drugs that exhibited significant inhibitory potential on cell proliferation with the HER2+ breast cancer lines. These include: mercaptopurine, nelfinavir mesylate, gefitinib, triciribine, and 6- α -methylprednisolone. It was reported that four of the drugs, excluding NFV, offered negligible to relatively good inhibitory activity, and were thus discarded as viable therapeutics. Shim *et al.* reported the action of NFV in several breast cancer cell lines on a collection of yeast strain and identified Hsp90 as its target. The possibility that NFV had an anticancer effect on yeast raised the possibility that Hsp90 is the target of nelfinavir in mammalian cells. Shim *et al.* further projected that since NFV has no effect on trypsin digestion of *N*- and middle domain of the Hsp90, it is likely that NFV binds to the Hsp90 C-terminal domain and induces conformational changes in the protein; which is a different mechanism from other Hsp90 inhibitors (20). A study performed by Peterson further affirm this assumption that the CTD, which is responsible for maintaining Hsp90 functional homodimeric state and coordinating interactions with several Hsp90 co-chaperones, show some promising binding mechanism (14).

A study presented by Srirangam *et al.* reported that another protease inhibitor, ritonavir, partially inhibit the functions of Hsp90 (21). Based on the findings from the studies performed by Shim *et al.* and Bernstein *et al.*, the FDA-approved HIV-1 protease inhibitors (PIs) showed to be promising in the inhibition of cancer, which is distinct from their ability to inhibit HIV protease (20, 22).

The study performed by Shim *et al.*, however, have some limitations. Foremost, the precise mode of interaction between NFV and Hsp90 remains a mystery. Secondly, the objectivity and interpretation of experimental outcome on mammalian cells was not stated. The probability that other PIs will be as potent as NFV is uncertain. The lack of a full X-ray crystal structure of the human Hsp90 might have been an important factor that significantly contributed to ambiguity of the binding mechanism of these drugs. From the lack of information with regards to binding modes, specific binding mechanism and conformation, prompted our extensive computational investigation in order to explore the exact binding modes of HIV PR inhibitors against the human Hsp90.

In this work, a homology model for the human Hsp90 was built. Comparative MD simulations and binding free energy calculations for nine FDA-approved HIV PR inhibitors (Figure 5.2) at the two different binding domains of Hsp90, C-terminal domain (CTD) and N-terminal domain (NTD), were performed. A total of 18 MD simulations protocol (5 ns each) were performed (section 2.4 holds more details on the MD simulation protocol) and post-dynamic analyses were also performed. The compilation of the computational and molecular modeling tools presented in this study could serve as powerful tools to understand protein structures and dynamics which could be incorporated in drug discovery, design, development, repositioning and optimization of potential Hsp90 inhibitors as anticancer agents.

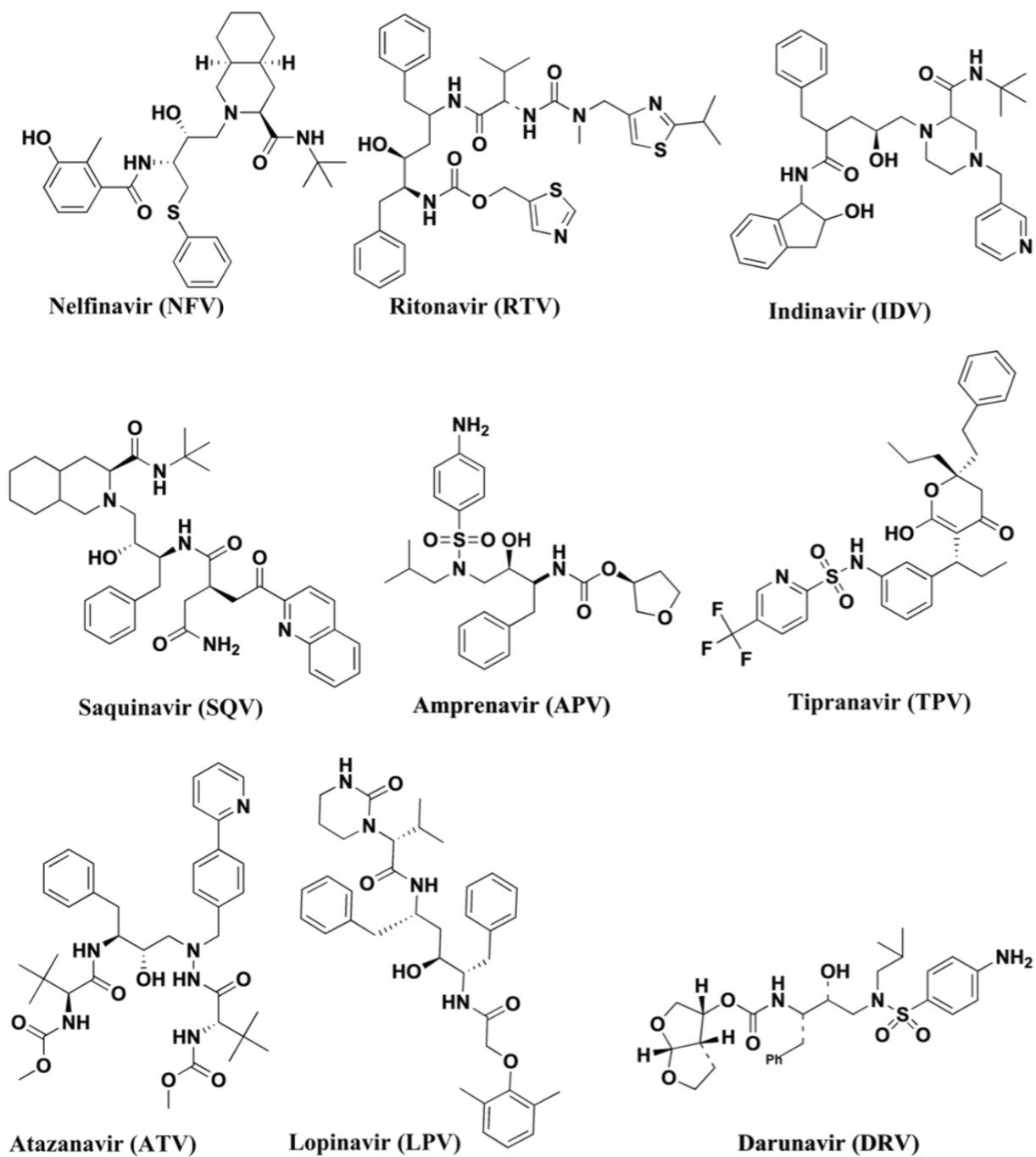


Figure 5.2. A schematic representation of 2D structures of the nine FDA-approved HIV-1 protease inhibitors.

2. Computational Methodology

2.1. Homology modeling of human Hsp90 protein structure

Due to the absence of a full human crystal structure of Hsp90 that comprises of both the NTD and CTD, the crystal structure of the human Hsp90 was modeled using the protein sequence obtained from Uniprot (Uniprot ID: P08238). The full human Hsp90 homologue was modeled by using three crystal structures of Hsp90 as templates: Hsp90 from *saccharomyces cerevisiae* (PDB Code: 2CG9), which contained the ATP bound in its active site; Hsp90 MD from *homo sapiens* (PDB Code: 3PRY) and Hsp90 CTD from *Leishmania major* (PDB Code: 3HJC). Homology modeling was performed using the Modeler Software version 9.1 (23) add-on in Chimera (24). Multiple sequence alignment was performed on CLUSTAW (25). The active site residues were determined using Chimera Multi-align Viewer and validated using the Site-Hound web program (26). The homology model of the human Hsp90 was energy minimized and equilibrated via molecular dynamics simulations (see section 2.4) and then used for subsequent simulations.

The sequence of the target protein was uploaded unto PSIPRED V3.3 (27, 28) in order to obtain a predicted 3D secondary structure of the enzyme. Comparing the homologue to the predicted 3D structure and assessment of the bond angles and torsional strain shows the validation of the homology model. A Ramachandran plot for the analyses of bond angles and torsional strain was generated using Maestro (29). MolProbity (30) result shows that 98% of all residues are in the favoured regions and >99.8% of all residues are in the allowed regions which leaves a list of 20 outliers. The list shows that none of the active site residues are part of these outliers (See Supplementary Material-S2).

2.2. Defining the active site residues in the Hsp90 homology model

Two binding sites (NTD and CTD) are known to exist in the Hsp90 based on previous reports (14, 31). The binding site of the N-terminal domain includes: Leu43, Asn46, Lys53, Ile91, Asp97, Met93, Asn101, Ser108, Gly109, Phe133 and Thr179 (50). On the other hand, the C-terminal active site residues were identified as Gln523, Val534, Ser535, Lys538, Thr595, Tyr596, Gly597, Trp598 and Met602 in reference to 2CG9 crystal structure (14). The positions of these active site residues were mapped in the corresponding human Hsp90 β

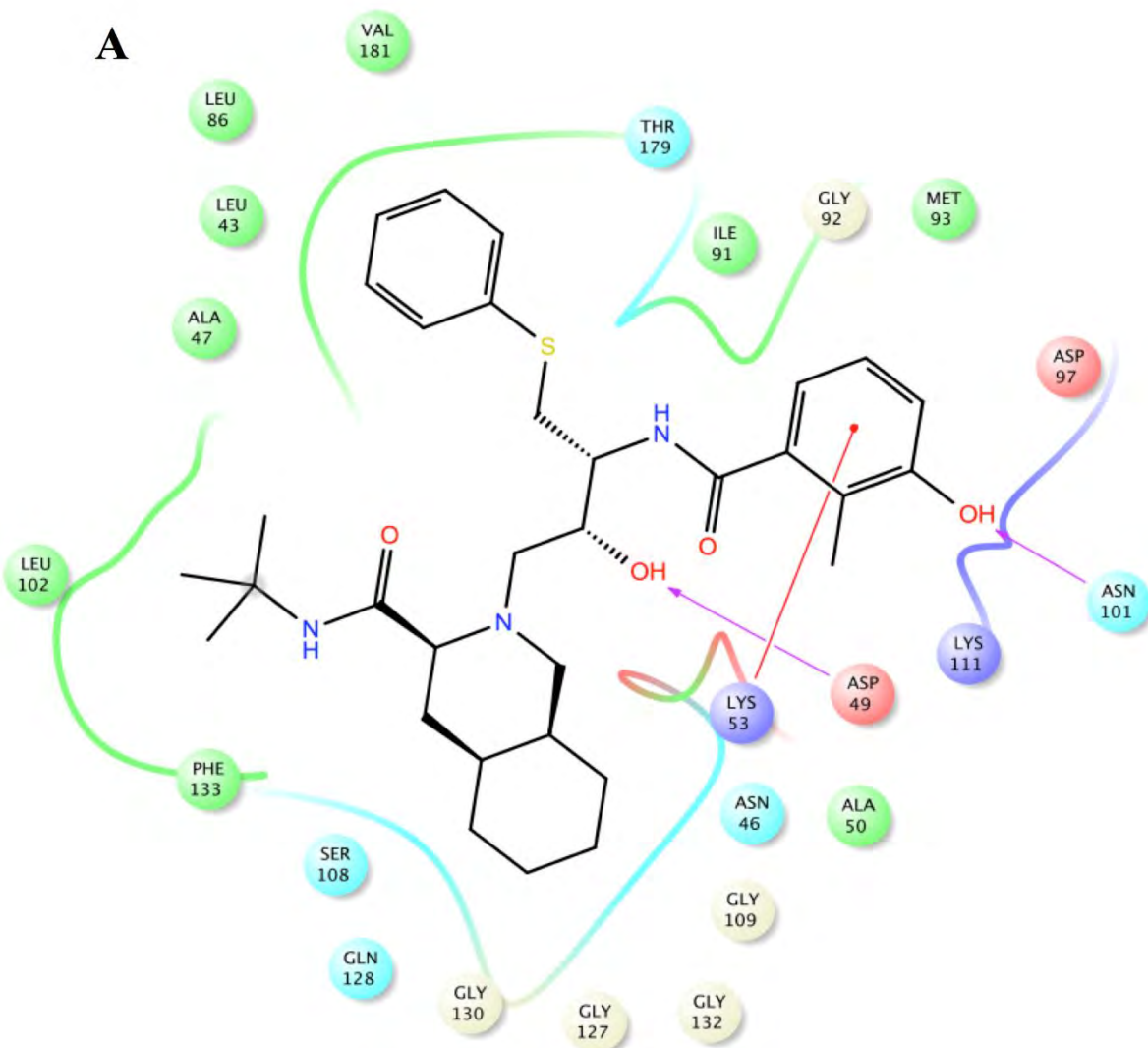
homologue to identify the NTD and CTD active site pockets for further docking and molecular dynamic simulations.

2.2.1. N-terminal Domain (NTD)

The Swiss-PDB Viewer (30) was used to align the human Hsp90 homologue within the crystal structure of the human Hsp90 NTD (PDB Code: 3PRY) that contain geldanamycin inhibitor where the active site residues were known (32, 33). Chimera software (34) was used to define the corresponding active site residues in the homology model. Figure 3 shows the human Hsp90-nelfinavir interaction at the NTD. The NTD active site residues were found to be Leu43, Asn46, Lys53, Ile91, Asp97, Met93, Asn101, Ser108, Gly109, Phe133 and Thr179.

2.2.2. C-terminal domain (CTD)

Due to the lack of information on the active site residues for the CTD, the active site residues were obtained from the Site-Hound web software (26). Closest active residue to the binding pocket, as shown in Figure 3, was selected and used for further modeling studies. The obtained active site residues for the human Hsp90 homologue CTD were Gln523, Val534, Ser535, Lys538, Thr595, Tyr596, Gly597, Trp598 and Met602.



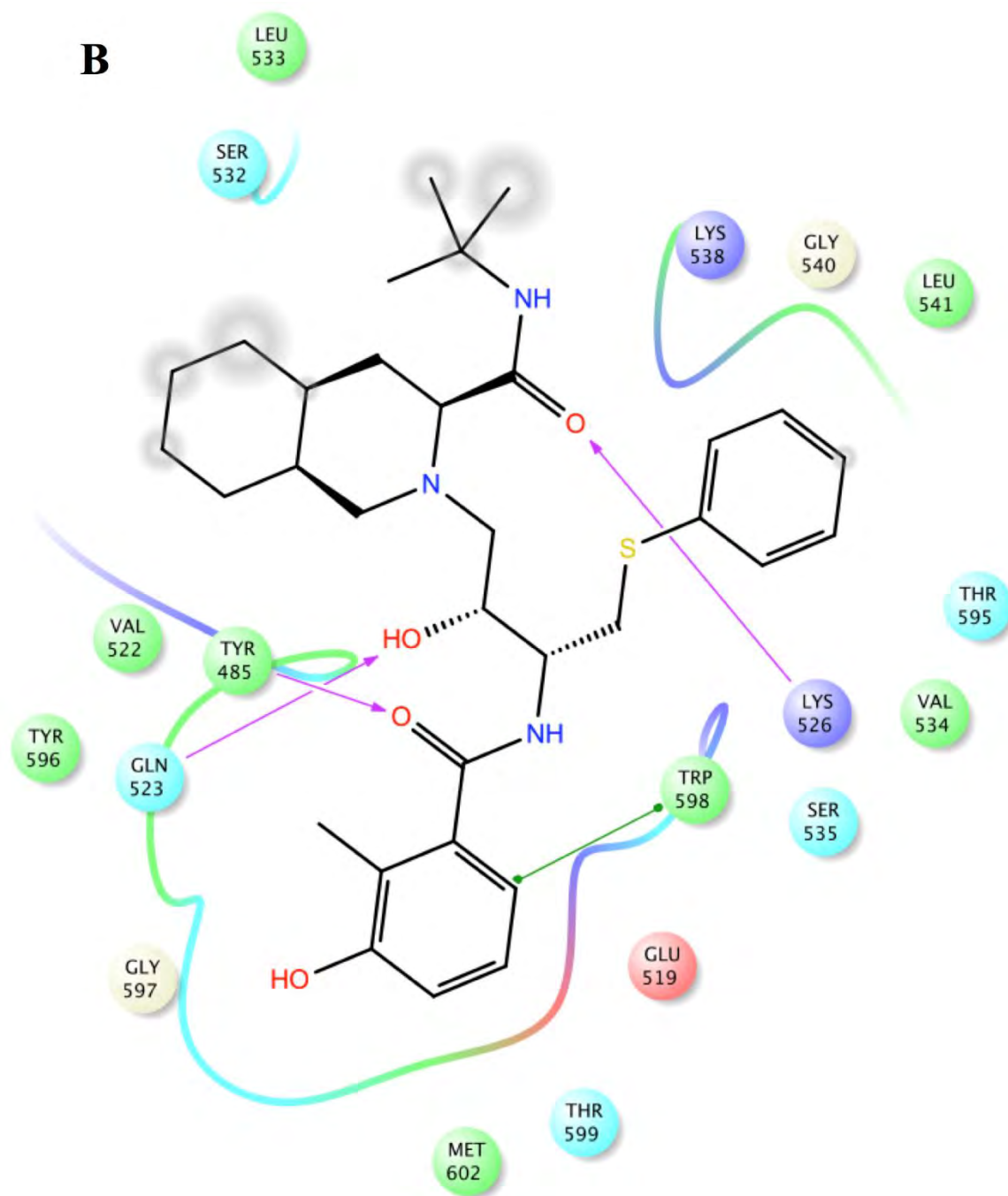


Figure 5.3. (A) Nelfinavir-human Hsp90 homologue interaction at the NTD (B) Nelfinavir-human Hsp90 homologue interaction at the CTD. Major contributions are from the residues that exhibit hydrophobic interactions (in green bubbles) and hydrogen bond interactions in purple arrows. Each illustration for other ligands at both terminals is provided in the supplementary material-S3.

2.3. Building Hsp90-HIV protease inhibitor complexes

In this study, AutoDock 4.0 suite was used as molecular docking tool in order to carry out the docking simulations. AutoDock has been found to be able to locate docking modes that are consistent with X-ray crystal structures (35, 36). AutoDock helps to simulate interactions between substrates or drug candidates as ligands and their macromolecular receptors of known three dimensional structures, allowing ligand flexibility described to a full extent elsewhere (37). In our docking simulations, the human Hsp90 homologue was used for performing docking. We opted to run virtual screening to assign the binding modes of HIV-1 protease inhibitors against Hsp90. To this end, 9 HIV-1 protease inhibitors were retrieved from ZINC database (38) and docked into the NTD and CTD active sites of Hsp90 homologue using Autodock Vina software (39). The grid box was defined with the grid parameters being X= 22, Y= 22 and Z= 20 for the dimensions and X= -126, Y= -33 and Z= 110 for the center grid box at the N-terminal domain; and grid parameters being X= 16, Y= 22 and Z= 28 for the dimensions and X= -81, Y= -54 and Z= 56 for the center grid box in the C-terminal domain. The human Hsp90-HIV-1 protease inhibitor complexes were obtained and used for all subsequent methods performed in this study.

2.4. Molecular Dynamics Simulations

Hsp90-HIV-1 protease inhibitor complexes were simulated using the GPU version of the PMEMD engine provided with the AMBER 12 package (40, 41). GAFF force field parameters for Hsp90-HIV-1 protease inhibitor complexes were calculated by antechamber module of AMBER 12 package. Hydrogen atoms of the proteins were added using the Leap module in AMBER 12 (41). The human Hsp90-HIV-1 protease complexes were obtained using Chimera and were solvated in an octahedron box of TIP3PBOX water with buffering distance of 8 Å between the protein surface and the box boundary (42); assuming normal charge states of ionizable groups corresponding to pH 7, sodium (Na⁺) counter-ions were added to achieve charge neutrality and to mimic biological environment more closely. Cubic periodic boundary conditions were imposed and the long-range electrostatic interactions were treated with the particle-mesh Ewald method implemented in AMBER 12 with a non-bonding cut-off distance of 10 Å. The partial atomic charges for the ligand were obtained using “antechamber” (43) module of AMBER 12. Initial energy minimization, with a restraint potential of 500 kcal/mol Å² applied to the solute, was carried out with the aid of the

SANDER module of the AMBER 12 program using the steepest descent method in AMBER 12 for 1000 steps followed by conjugate gradient protocol for 2000 steps.

Due to the lack of parameters needed for the ligand in the Cornell *et al.* force field (44), the missing parameters were created. Optimization of the ligands are first performed at the HF/6-31G* level with the Gaussian 03 package(45). The standard AMBER force field for bioorganic systems (ff03) (46) was used to define the topology and parameter files for the enzyme and protease inhibitors using “gaff” (47) based on the atom types of the force field model developed by Cornell *et al.* (40). The entire system was freely minimized for 1000 iterations. Heating was performed for 50ps from 0 to 300 K with harmonic restraints of 5 kcal/mol Å² using a Langevin thermostat with a coupling coefficient of 1/ps. The entire system was then equilibrated at 300 K with a 2fs time step in the NPT ensemble for 500 ps and Berendsen temperature coupling (48) was used to maintain a constant pressure at 1 bar. The SHAKE algorithm (49) was employed on all atoms so as to constrain the bonds of all hydrogen atoms. With no restraints imposed, a production run was performed for 5 ns in an isothermal isobaric (NPT) ensemble using a Berendsen barostat with a target pressure of 1 bar and a pressure-coupling constant of 2ps for analysis of the energy stabilization and RMSD values of the complex. The coordinate file was saved every 1ps and the trajectory was analyzed every 1ps using the Ptraj module implemented in AMBER 12.

2.5. Thermodynamic Calculation

The free binding energy of protease inhibitors (PIs) to the human Hsp90 homologue active site was analyzed by the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method (50-53). A single trajectory approach was used with 5000 snapshots at 50 ps interval of each simulation. From each snapshot, free binding energy (ΔG_{bind}) was computed from the following:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \quad (1)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S \quad (2)$$

$$E_{gas} = E_{int} + E_{vdw} + E_{ele} \quad (3)$$

$$G_{sol} = G_{GB} + G_{SA} \quad (4)$$

$$G_{SA} = \gamma SASA \quad (5)$$

Where E_{gas} is the gas-phase energy; E_{int} is the internal energy; E_{ele} and E_{vdW} are the Coulomb and van der Waals energies, respectively. E_{gas} was calculated using the ff03 force field. G_{sol} is the solvation free energy and can be decomposed into polar and nonpolar contributions. G_{GB} is the polar solvation contribution calculated by solving the GB equation. G_{SA} is the nonpolar solvation contribution estimated by the solvent accessible surface area (SASA) and was determined using a water probe radius of 1.4 Å. The surface tension constant γ was set to 0.0072 kcal/(mol·Å²) (54). The T and S are the temperature and the total solute entropy, respectively. S was calculated by classical statistical thermodynamics, using normal mode analysis (55). Normal mode analysis was carried out in the AMBER 12 normal mode (NMODE) module. Due to the high computational cost in the entropy calculation, 100 snapshots were extracted from the last equilibrated 5 ns trajectory of the simulation with 50 ps time intervals.

To theoretically evaluate the reliability of the calculated ΔG values, the standard error (SE) of the calculated free binding energy was estimated by using equation 6, which is related to the number (N) of snapshots chosen for the calculations (37).

$$SE = RMSF / N \quad (6)$$

3. Results and Discussion

3.1. The Human Hsp90 homology model

In an effort to work with a human Hsp90 that comprised of the middle domain (MD), C-terminal domain (CTD) and N-terminal domain (NTD), a homology model of the human Hsp90 β was generated. Using the Hsp90 from *Saccharomyces cerevisiae* (PDB Code: 2CG9), crystal structure of Hsp90 MD from *Homo sapiens* (PDB Code: 3PRY) and a crystal structure of Hsp90 CTD from *Leishmania major* (PDB Code: 3HJC) as structural templates, an human Hsp90 homologue was constructed using software “Modeller”9.1 (23). Structural similarity between the three proteins showed good identity in and around the active site, NTD, MD and CTD, with the majority of active site residues having nearly identical positions (Figure 5.4a). The sequences shared a 64.11% (2CG9), 97.39% (3PRY) and 60.14% (3HJC) similarity according to the Multi-align Viewer tool in Chimera, and the model had a zDOPE score of 0.25 after modeling (Figure 5.4b).



Figure 5.4. (A) Superimposed structures of the 2CG9 (gold), 3PRY (purple) and 3HGC (blue) and the generated human homologue sequence (green), (B) Generated homology model of the human Hsp90 β .

As docking is the most appropriate tool to provide better description of the binding theme of inhibitors (56, 57), molecular docking was performed to elucidate the binding mode of the human Hsp90 homologue and its inhibitors. To this end, docking of HIV-1 protease inhibitors was successfully undertaken giving an idea on how these inhibitors are positioned on the binding pocket on the NTD and CTD. Figure 5.5 shows docked structure of the human Hsp90 homologue in complex with NFV. Nine FDA-approved HIV-1 inhibitors were docked to the human Hsp90 allowing for the construction of enzyme-ligand complexes. Interestingly, all the HIV-1 PIs showed better binding to the active site residues of the human Hsp90 homologue and were perfectly situated on the binding pocket of the human Hsp90 homologue NTD and CTD as how the known inhibitors of Hsp90 bind (33).

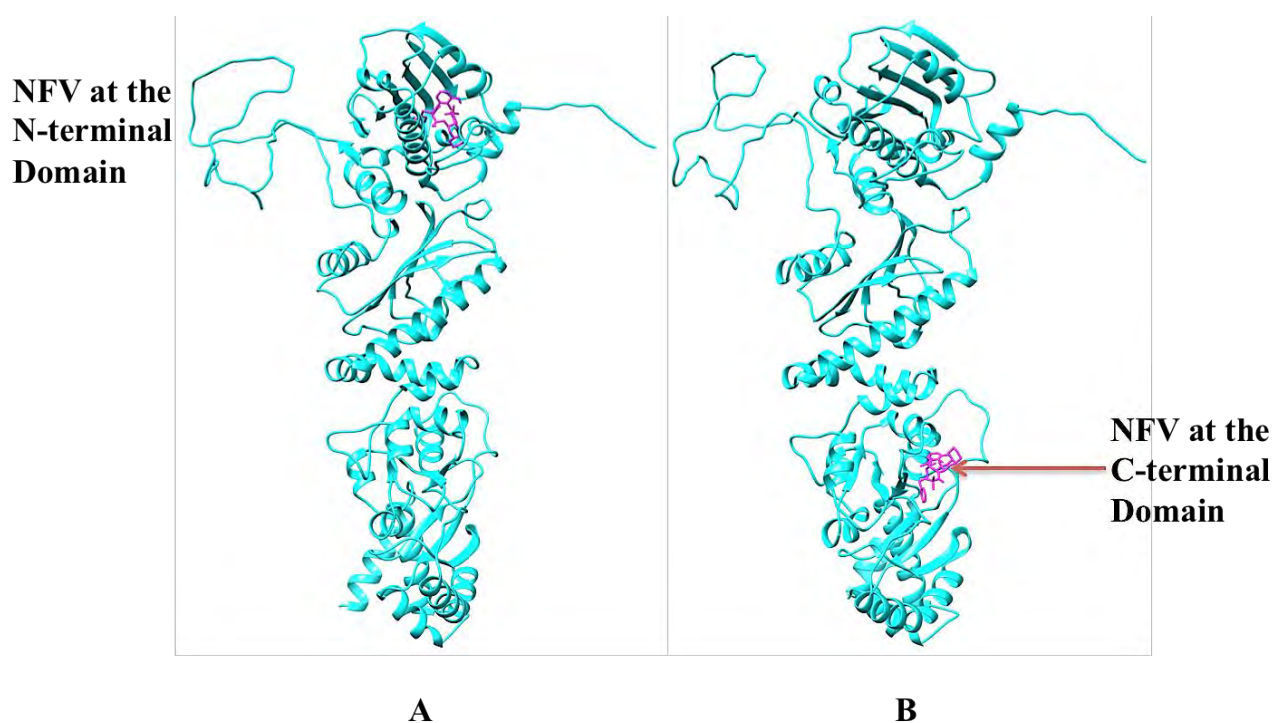


Figure 5.5. The human Hsp90 homologue docked with NFV at (a) N-terminal domain and (b) C-terminal domain.

3.2. Molecular Dynamics (MD) simulations

The structural dynamics of all HIV-1 PIs, which are bound to the human Hsp90 homologue were analysed by performing 5 ns molecular dynamics simulation (58-60) and free binding energy calculations (61, 62), which have proved to be very useful and successful in understanding the molecular basis of drug inhibition to different biological targets. It provides useful structural and energetic information about the interaction between the inhibitors and the targets. The system stability and overall convergence of simulations were monitored in terms of Root Mean-Square Deviation (RMSD), and potential energy of the protein backbone atoms. The RMSD and potential energy enabled us to verify that equilibration was achieved.

3.2.1. Post-dynamic analyses:

Stability of Molecular Dynamics simulation

As shown in Figure 5.6, the system was well equilibrated where the RMSD value of all nine complexes at the NTD and CTD did not exceed 2Å. Stabilisation of the RMSD was achieved after 1000 ps and 1800 ps, for NTD and CTD respectively. The fluctuations of potential

energies at the NTD were <2000 kcal/mol for the duration of 5 ns whereas the CTD had fluctuations in the potential energy of ≤ 8000 kcal/mol. Overall, it was observed that equilibration at NTD and CTD was achieved although they occurred at different time frames. The potential energy observed for the inhibitors at the CTD indicate that there is an initial energy increase that is due to the heating up of the system. A decrease in potential energy was observed followed by stabilization, implying that the system folded up to a state more stable than the starting linear structure. This trend was observed for all complexes at the CTD.

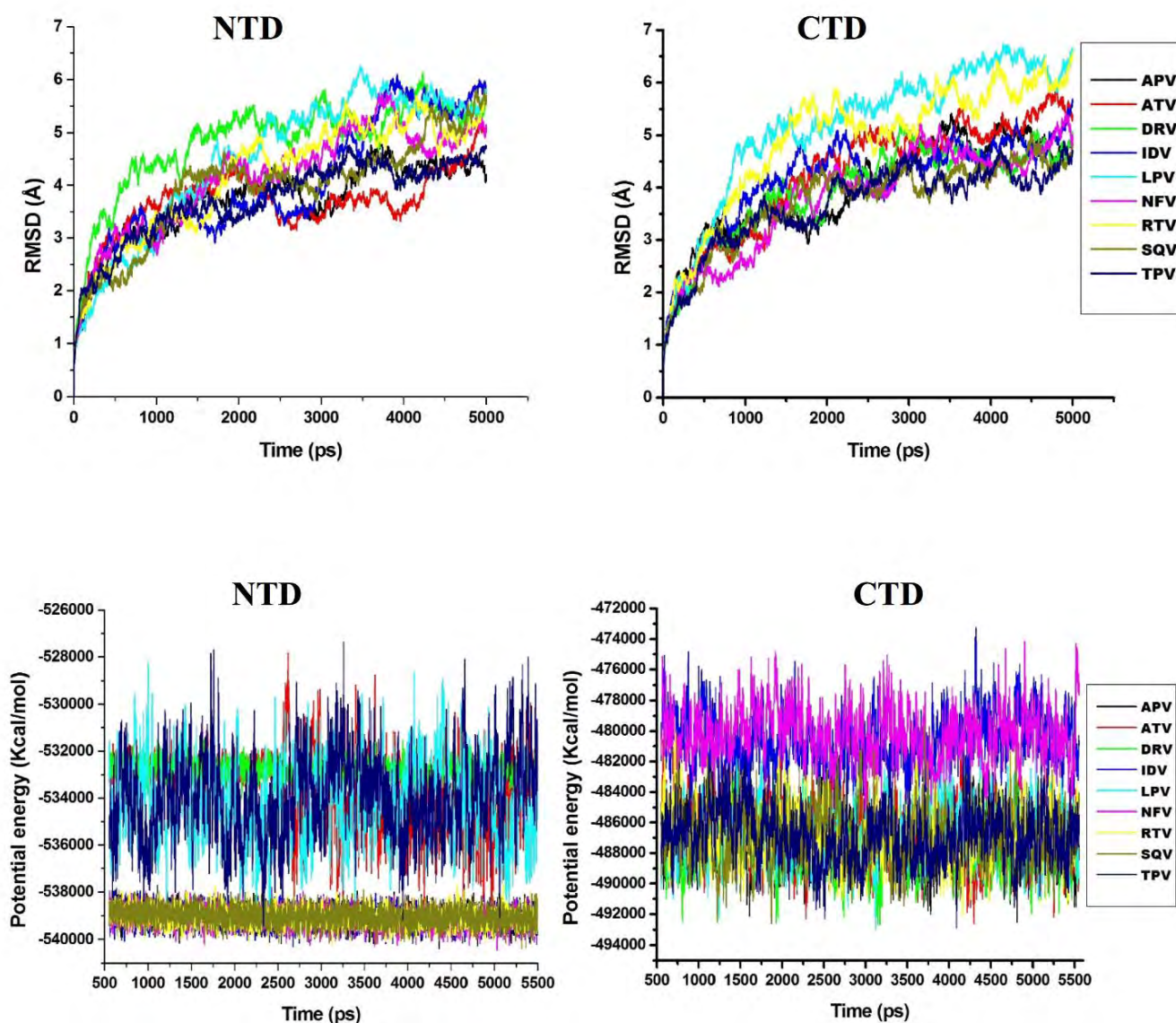


Figure 5.6. Comparative RMSD and potential energy plot of all ligands binding at the N-terminal domain and C-terminal domain respectively. (Individual plot for each ligand is provided in the supplementary material).

RMSF calculations

The root mean square fluctuations (RMSF) of the average structure from the trajectory report the residue and atomic fluctuations to check if the simulation results are in accordance with the crystal structure. Figure 5.7 shows the residue based RMSF of the NTD and CTD simulations. The amino acid residues in the region of 216-287 show higher fluctuation in both the NTD and CTD in comparison to the other amino acid residues which all exhibit lower fluctuations. It was observed that the amino acids in the region of 216-287 demonstrated no significant effect on how the human Hsp90 homologue interacted with the HIV-PIs at both terminals.

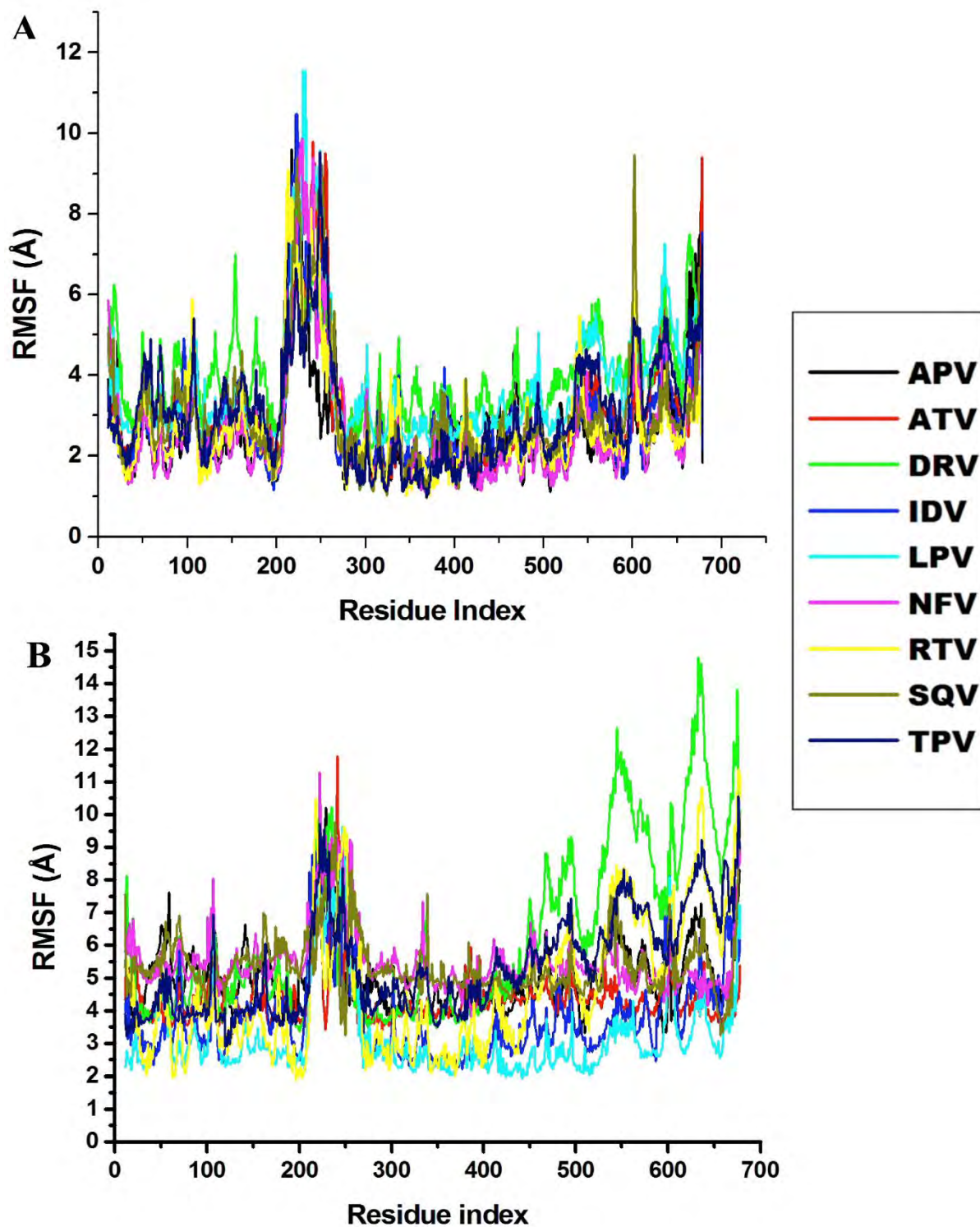


Figure 5.7. (A) Comparative RMSF plot of all ligands binding at the N-terminal domain and (B) Comparative RMSF plot of all ligands binding at the C-terminal domain. (Individual plot for each ligand is provided under the supplementary material- Appendix 1).

3.2.2. MM/GBSA free binding energy calculations

In the present study, we applied MM/GBSA as it has been demonstrated to be more efficient for protein-drug systems (63-65), carbohydrates (66) and nucleic acids (67). It utilizes a fully pairwise potential that is useful for decomposing the total binding free energy into atomic or group contributions (68). In addition, the MM/GBSA was successfully applied for ligand binding interactions with HIV PR with multidrug resistance (69-71). All components of molecular mechanics, solvation energy and free binding energies for the binding of the PIs (ΔG_{bind} in kcal/mol) were calculated from the MD trajectories using the MM/GBSA method (50-53) implemented in AMBER 12 (41). This method explores the type of established interactions, calculating separately the components of the internal energy, the interaction energy and the free energy of Gibbs of solvation (72). 100 snapshots were extracted at a time interval of 10 ps from the 5 ns of MD trajectories for the analysis of the binding free energy. MD simulations were used to examine the solvation of the binding pockets of the human Hsp90 homologue (NTD and CTD) in complex with the nine PIs. The values of the free binding energy presented in Table 5.1 show that for all the PIs, there is a higher binding in the NTD as compared to the CTD. The free binding energy of NFV being -83.03 kcal/mol at the NTD and -39.3 kcal/mol at the CTD. This value shows a significant difference (~ 43.73 kcal/mol) between the interaction energy at the NTD and CTD. Indinavir (IDV) on the other hand, gives the high value of the free binding energy of -67.3 kcal/mol, with an internal Van der Waal (ΔG_{vdw}) energy of -81.3 kcal/mol at the NTD whereas there is a huge drop of the free binding energy (-9.7 kcal/mol) and van der Waals energy (-41.8 kcal/mol) at the CTD. Such a great reduction in binding affinity (~ 57.6 kcal/mol) indicates relatively weaker ligand binding at the CTD.

The major favourable contributions observed for the human Hsp90 homologue-PIs complexes free binding energies occur at the NTD which are significantly higher compared to the CTD. This trend is observed for all complexes except RTV which has a reasonably good free binding energy (-64.8 kcal/mol) at the NTD and (-63.9 kcal/mol) at the CTD with a difference of (~ 0.9 kcal/mol). There appears to be a probability that RTV has a good efficacy at both terminals and should therefore be considered for experimental evaluation.

Table 5.1. Binding free energies of the FDA-approved protease inhibitors against the human Hsp90 homologue.

HIV-1 PIs Inhibitors	Contributions ^a						
	ΔE_{ele}	ΔE_{vdW}	ΔG_{SA}	ΔG_{GB}	ΔG_{gas}^b	ΔG_{sol}^c	ΔG_{bind}^d
NFV	-83.23±1.3 -103.0±0.9	-57.7±0.1 -51.9±0.9	-7.72±0.1 -7.3±0.1	100.84±1.4 128.9±1.2	-25.52±1.3 -160.9±1.7	108.56±1.4 121.7±1.1	-83.03±0.5 -39.3±0.6
RTV	-14.9±0.3 -17.1±0.4	-80.1±0.3 -76.2±0.4	-10.7±0.02 -9.2±0.04	40.8±0.3 38.6±0.4	-94.9±0.5 -93.3±0.9	30.1±0.3 29.4±0.4	-64.8±0.3 -63.9±0.4
SQV	-273.3±1.0 -97.7±1.9	-89.5±0.3 -38.4±1.0	-10.7±0.01 -6.2±0.1	308.4±0.9 119.3±2.1	-362.8±1.1 -136.1±2.8	297.7±0.9 113.1±1.9	-64.8±0.4 -22.9±0.9
IDV	-652.2±1.6 -266.0±1.6	-81.3±0.4 -41.8±0.2	-10.1±0.02 -5.4±0.05	676.2±1.5 303.6±1.6	-733.4±1.6 -307.8±1.7	666.1±1.5 298.1±1.6	-67.3±0.7 -9.7±0.4
APV	-4.9±0.6 -7.9±0.3	-67.5±1.7 -48.9±0.5	-7.5±0.2 -6.4±0.04	25.2±0.8 22.2±0.3	-72.4±1.9 -56.9±0.6	17.7±0.7 15.8±0.3	-54.7±1.4 -41.1±0.5
LPV	-8.0±0.7 -12.2±0.3	-83.3±0.4 -56.5±0.4	-9.7±0.04 -7.2±0.05	30.7±0.6 32.3±0.3	-91.4±0.8 -68.8±0.5	21.1±0.6 25.1±0.3	-70.3±0.5 -43.7±0.4
ATV	-7.4±0.4 -13.9±0.3	-89.8±0.4 -54.8±0.4	-11.1±0.03 -7.0±0.04	36.8±0.4 36.3±0.3	-97.2±0.6 -68.7±0.5	22.8±0.4 29.3±0.3	-74.4±0.4 -39.4±0.4
DRV	-16.9±0.5 -12.9±0.3	-75.1±0.3 -47.1±0.3	-9.2±0.03 -6.2±0.4	36.8±0.5 29.0±0.3	-92.1±0.7 -60.0±0.4	27.6±0.5 22.9±0.3	-64.5±0.5 -37.1±0.2
TPV	-11.0±0.3 -16.1±0.3	-77.6±0.3 -51.5±0.32	-9.7±0.03 -6.7±0.03	36.4±0.3 35.5±0.3	-88.6±0.41 -67.7±0.5	26.7±0.3 28.8±0.3	-61.9±0.3 -38.9±0.3

^a All energies of the NTD and CTD, with corresponding standard errors, are in kcal/mol.

^b $\Delta E_{gas} = \Delta E_{int} + \Delta E_{vdW} + \Delta E_{ele}$

^c $\Delta G_{sol} = \Delta G_{SA} + \Delta G_{GB}$

^d $\Delta G_{bind} = \Delta G_{gas} + \Delta G_{sol} - T\Delta S$

Shim *et al.* (2012) analyzed the activities of some drugs, which include mercaptopurine, nelfinavir mesylate, gefitinib, triciribine and 6- α -methylprednisolone, which showed least to relatively good selectivity for HER2+ breast cancer lines. The action of nelfinavir in HER2+ breast cancer cells was tested by screening a collection of yeast strains and acknowledged the Hsp90 protein as a potential target. In the study conducted by Shim, nelfinavir strongly inhibited certain HER2 signaling events and also had the highest potency (IC₅₀ 10 μ M), which suggests that it may be effective in HER2+ breast cancer patients with the same dosage regimen administered to HIV-1-infected patients (20). Based on Shim *et al.* study, it is clear that PIs inhibit a wide variety of malignant cell lines with nelfinavir having the best inhibition profile. Despite extensive experimental studies performed on the anticancer activity of NFV,

the precise molecular and binding mechanism responsible for its inhibitory effect on either yeast or human Hsp90 remains anonymous.

From Table 5.1, it is evident that nelfinavir binds readily to the human Hsp90 homologue NTD ($\Delta G_{\text{bind}} -83.23$ kcal/mol) which is in accordance with the experimental findings from the study conducted by Shim *et al.* (20). This finding implies that the simulation protocol of this work is appropriate and reliable.

In comparison to the CTD, the human Hsp90 NTD binding pocket was found more promising than the CTD binding pocket. According to the results of the ΔE_{vdw} calculation, the NTD domain also showed higher van der Waals interaction energies (ranging from -57.7 to -89.8 kcal/mol), whereas the CTD showed lower interaction energies (ranging from -38.4 to -76.2 kcal/mol) as compared to the NTD. It was observed that other PIs showed better interaction to the NTD implying that the NTD might be the preferred domain for binding of PIs to the human Hsp90. The study performed by Shim *et al.*, showed that nelfinavir inhibit Hsp90 at the site distinct from that of the known inhibitor (20). At this point, more studies is required to further verify the binding mode of the human Hsp90 CTD to HIV-1 PIs.

3.2.3. Per-residue interaction energy decomposition analysis

The binding free energy was further decomposed into contributions from each human Hsp90 β amino acid residue. It can be observed from the energy decomposition analysis at the NTD that the amino acid residues with major contributions were: Leu43, Asn46, Lys53, Ile91, Asn101, Ser108, Gly109, Met93, Phe133, Thr179 and those with minor contributions towards the interaction energy were residues Asp97, Val131, Val145, Val181 and Tyr134. At the CTD, the amino acid residues with major contributions were Gln523, Val534, Lys538, Ser535, Thr595, Tyr596, Trp598 and Met602 and those with minor contributions were Tyr485, Glu519, Val522, Gln524, Lys526, Leu533, Val534, Thr599 and Tyr619. These contributions can be due to their hydrophobicity nature as shown in Figure 5.8. These trends were maintained in all complexes at both terminals.

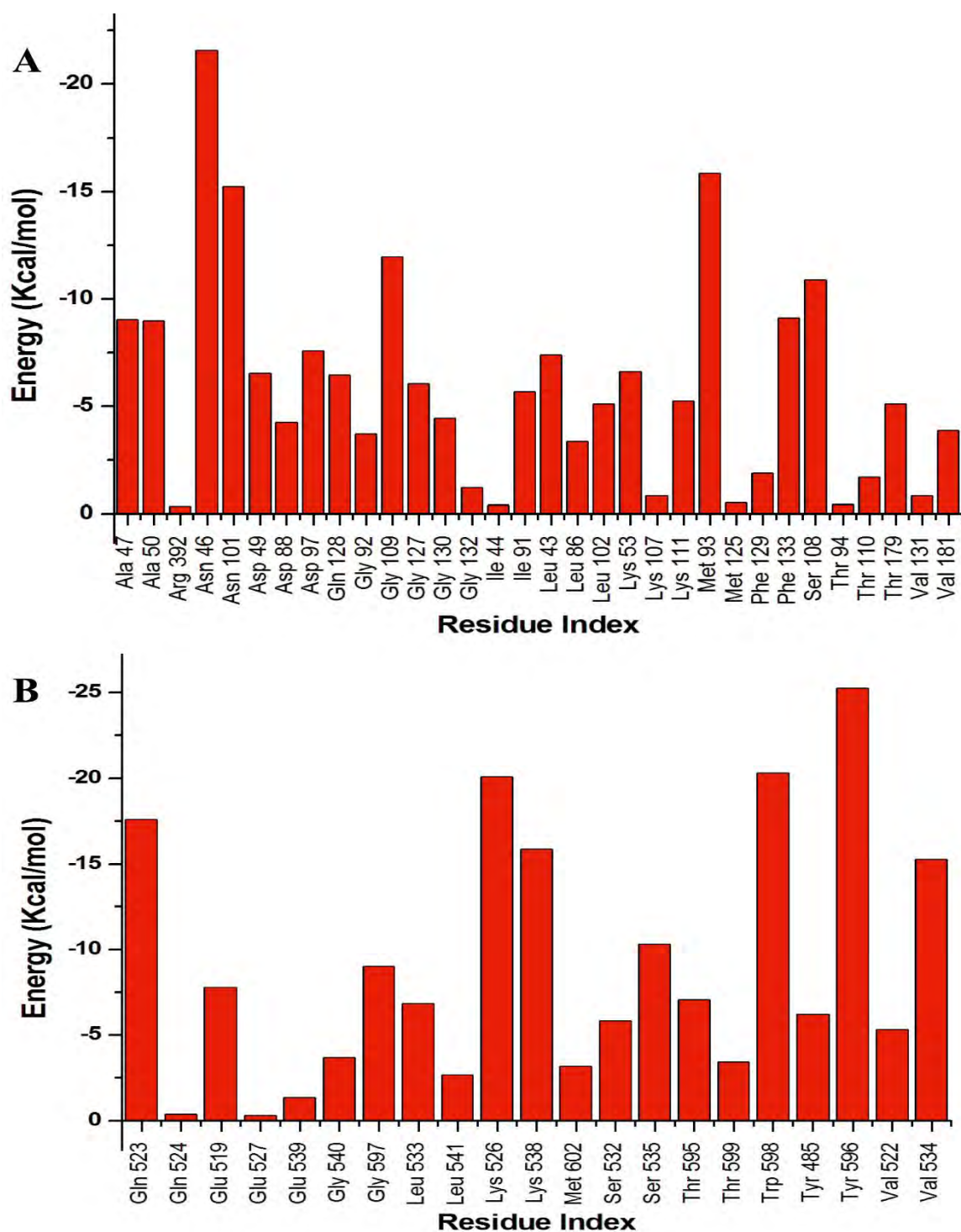


Figure 5.8. (A) Residues that contributed to the human Hsp90-nelfinavir binding at the NTD and (B) Residues contributing to the human Hsp90-nelfinavir binding at the CTD. (Illustrations for other ligands are provided in the supplementary material-S4).

4. Conclusion

In this study, the potential mechanism of binding of nine PIs to the human Hsp90 β homologue at the NTD and CTD was investigated. Some active site residues have been identified in both terminals in previous studies and this brought about the investigation and validation of these theories by computational methods. These methods include: Virtual screening, Molecular dynamics (MD) simulation, binding free energy calculation using MM/GBSA method implemented in AMBER 12, potential energy, per-residue energy decomposition, RMSF and RMSD calculations.

From the above investigation of the CTD and NTD of the human Hsp90 homologue binding to HIV-1 protease inhibitors, it can be concluded that these HIV-1 PIs interact better at the NTD than at the CTD. However, ritonavir has promising binding affinities at both terminals and should therefore be considered for experimental evaluation. Additionally, apart from ritonavir, saquinavir and nelfinavir; further investigation needs to be carried out on the CTD to look for more promising drugs that could bind to its active site residues. As at the time of this study, there was limited experimental information available regarding this domain. It was reported in previous studies that nelfinavir was found to be more potent than other PIs tested against breast line cell cancer, however, computationally, result shows that nelfinavir as well as other PIs may also have good binding affinity, thus, excellent candidates for further evaluation as anticancer agents.

Supplementary Materials

RMSF vs. time, RMSD vs. time and Potential Energy (kcal/mol) vs. time data for the human Hsp90 homologue in complex with HIV-1 PIs at both NTD and CTD, Per-residue decomposition at NTD and CTD; ligand-complex interaction at NTD and CTD; Molprobit result including all validation data are provided in the supplementary material.

Acknowledgement

The authors acknowledge the School of Health Sciences, UKZN, for financial support and CHPC for technical support.

Conflicts of Interest

Authors declare no potential conflicts of interest.

References

1. World Health Organization. (2008) The global burden of disease: 2004 Update, In http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf, 18 Jan. 2014.
2. World Health Organization. (2014) World cancer burden (2012), Cancer Research UK.
3. American Cancer Society. (2014) Cancer Facts & Figures 2014, 2014 ed., Atlanta: American Cancer Society.
4. Jemal, A., Bray, F., Forman, D., O'Brien, M., Ferlay, J., Center, M., and Parkin, D. M. (2012) Cancer burden in Africa and opportunities for prevention, *Cancer* 118, 4372-4384.
5. Unnati, S., Ripal, S., Sanjeev, A., and Niyati, A. (2013) Novel anticancer agents from plant sources, *Chin. J. Nat. Med.* 11, 16-23.
6. Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008, *Int. J. Cancer* 127, 2893-2917.
7. Cos, S., González, A., Martínez-Campa, C., Mediavilla, M. D., Alonso-González, C., and Sánchez-Barceló, E. J. (2006) Estrogen-signaling pathway: A link between breast cancer and melatonin oncostatic actions, *Cancer Detect. Prev.* 30, 118-128.
8. Stern, D. F. (2000) Tyrosine kinase signalling in breast cancer - ErbB family receptor tyrosine kinases, *Breast Cancer Res.* 2, 176-183.
9. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) Human-breast cancer - correlation of relapse and survival with amplification of the HER-2 neu oncogene, *Sci.* 235, 177-182.
10. Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., Burchmore, M., Shak, S., Stewart, S. J., and Press, M. (2002) Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer, *J. Clin. Oncol.* 20, 719-726.
11. Gajria, D., and Chandarlapaty, S. (2011) HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies, *Expert Rev. Anticancer Ther.* 11, 263-275.
12. Kurebayashi, J. (2001) Biological and clinical significance of HER2 overexpression in breast cancer, *Breast Cancer* 8, 45-51.
13. Grover, A., Shandilya, A., Agrawal, V., Pratik, P., Bhasme, D., Bisaria, V., and Sundar, D. (2011) Hsp90/Cdc37 Chaperone/co-chaperone complex, a novel junction anticancer target elucidated by the mode of action of herbal drug Withaferin A, *BMC Bioinformatics* 12, 1-13.
14. Peterson, L. B. (2012) Investigation of the Hsp90 C-terminal binding site, Novel inhibitors and isoform-dependent client proteins, In https://kuscholarworks.ku.edu/dspace/bitstream/1808/10218/1/Peterson_ku_0099D_12224_DATA_1.pdf, 20 Aug. 2013.
15. Dollins, D. E., Warren, J. J., Immormino, R. M., and Gewirth, D. T. (2007) Structures of GRP94-Nucleotide complexes reveal mechanistic differences between the hsp90 chaperones, *Mol. Cell* 28, 41-56.
16. Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1997) Identification and structural characterization of the ATP/ADP-Binding site in the Hsp90 molecular chaperone, *Cell* 90, 65-75.

17. Ashburn, T. T., and Thor, K. B. (2004) Drug repositioning: Identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov.* 3, 673-683.
18. Gills, J., Lo Piccolo, J., Tsurutani, J., Shoemaker, R. H., Best, C. J. M., Abu-Asab, M. S., Borojerdi, J., Warfel, N. A., Gardner, E. R., Danish, M., Hollander, M. C., Kawabata, S., Tsokos, M., Figga, W. D., Steeg, P. S., and Dennis, P. A. (2007) Nelfinavir, a lead HIV protease inhibitor, is a broad-spectrum, anticancer agent that induces endoplasmic reticulum stress, autophagy, and apoptosis in vitro and in vivo, *Clin. Cancer Res.* 13, 5183-5194.
19. Osborne, C. K. (1998) Steroid hormone receptors in breast cancer management, *Breast Cancer Res. Treat.* 51, 227-238.
20. Shim, J. S., Rao, R., Beebe, K., Neckers, L., Han, I., Nahta, R., and Liu, J. O. (2012) Selective Inhibition of HER2-Positive Breast Cancer Cells by the HIV Protease Inhibitor Nelfinavir, *J. Natl. Cancer Inst.* 104, 1576-1590.
21. Srirangam, A., Mitra, R., Wang, M., Gorski, J. C., Badve, S., Baldrige, L., Hamilton, J., Kishimoto, H., Hawes, J., Li, L., Orschell, C. M., Srour, E. F., Blum, J. S., Donner, D., Sledge, G. W., Nakshatri, H., and Potter, D. A. (2006) Effects of HIV protease inhibitor ritonavir on Akt-regulated cell proliferation in breast cancer, *Clin. Cancer Res.* 12, 1883-1896.
22. Bernstein, W. B., and Dennis, P. A. (2008) Repositioning HIV protease inhibitors as cancer therapeutics, *Curr Opin HIV AIDS* 3, 666-675.
23. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. y., Pieper, U., and Sali, A. (2001) Comparative protein structure modeling using MODELLER, *Curr. Protoc. Protein Sci.*
24. Eric F. Pettersen, T. D. G., Conrad C. Huang, Gregory S. Couch,, and Daniel M. Greenblatt, E. C. M., Thomas E. Ferrin. (2004) UCSF chimera—A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25, 1605-1612.
25. CLUSTAW. <http://www.genome.jp/tools/clustalw/>, 16 Oct. 2013.
26. Sitehound-web. <http://scbx.mssm.edu/sitehound/sitehound-web/Input.html>, 15 Oct. 2013.
27. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* 292, 195-202.
28. Buchan, D. W. A., Ward, S. M., Lobley, A. E., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2010) Protein annotation and modelling servers at University College London, *Nucleic Acids Res.* 38, W563-W568.
29. Schrodinger. <http://www.schrodinger.com/>, 20 Oct. 2013.
30. Swiss PDB. <http://www.spdbv.vital-it.ch/disclaim.html>, 21 Oct. 2013.
31. Kim, Y. S., Alarcon, S. V., Lee, S., Lee, M. J., Giaccone, G., Neckers, L., and Trepel, J. B. (2009) Update on Hsp90 Inhibitors in Clinical Trial, *Curr. Top. Med. Chem.* 9, 1479-1492.
32. Jez, J. M., Chen, J. C. H., Rastelli, G., Stroud, R. M., and Santi, D. V. (2003) Crystal structure and molecular modeling of 17-DMAG in complex with human Hsp90, *Chem. Biol.* 10, 361-368.
33. Stebbins, C. E., Russo, A. A., Schneider, C., Rosen, N., Hartl, F. U., and Pavletich, N. P. (1997) Crystal structure of an Hsp90-geldanamycin complex: Targeting of a protein chaperone by an antitumor agent, *Cell* 89, 239-250.
34. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25, 1605-1612.
35. Dym, O., Xenarios, I., Ke, H. M., and Colicelli, J. (2002) Molecular docking of competitive phosphodiesterase inhibitors, *Mol. Pharmacol.* 61, 20-25.

36. Rao, M. S., and Olson, A. J. (1999) Modelling of Factor Xa-inhibitor complexes: A computational flexible docking approach, *Proteins* 34, 173-183.
37. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* 19, 1639-1662.
38. Zinc database. <http://zinc.docking.org>, 18 Sep. 2013.
39. Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *J. Comput. Chem.* 31, 455-461.
40. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995), *J. Am. Chem. Soc.* 118, 2309-2309.
41. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The amber biomolecular simulation programs, *J. Comput. Chem.* 26, 1668-1688.
42. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79, 926-935.
43. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method, *J. Comput. Chem.* 21, 132-146.
44. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, *J. Am. Chem. Soc.* 117, 5179-5197.
45. Frisch MJ, T. G., Schlegel HB, Scuseria GE, Robb MA,, Cheeseman JR, M. J., Vreven T, Kudin KN, Burant, JC, M. J., Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada, M, E. M., Toyota K, Fukuda R, Hasegawa J, Ishida M,, Nakajima T, H. Y., Kitao O, Nakai H, Klene M, Li X, Knox, JE, H. H., Cross JB, Bakken V, Adamo C, Jaramillo J,, Gomperts R, S. R., Yazyev O, Austin AJ, Cammi R,, Pomelli C, O. J., Ayala PY, Morokuma K, Voth GA,, Salvador P, D. J., Zakrzewski VG, Dapprich S, Daniels, AD, S. M., Farkas O, Malick DK, Rabuck AD, Raghavachari, K, F. J., Ortiz JV, Cui Q, Baboul AG, Clifford S,, Cioslowski J, S. B., Liu G, Liashenko A, Piskorz P,, Komaromi I, M. R., Fox DJ, Keith T, Al-Laham MA, Peng, CY, N. A., Challacombe M, Gill PMW, Johnson B,, and Chen W, W. M., Gonzalez C, Pople JA. (2004) http://www.gaussian.com/g_misc/g03/citation_g03.htm, *Gaussian Inc, Wallingford, CT*.
46. 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., and Kollman, P. A. (2003) point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.* 24, 1999–2012.
47. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J. Comput. Chem.* 25, 1157-1174.
48. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath, *J. Chem. Phys.* 81, 3684-3690.

49. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes, *J. Comput. Phys.* *23*, 327-341.
50. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Acc. Chem. Res.* *33*, 889-897.
51. Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspect. Drug Discovery Des.* *18*, 113-135.
52. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized born solvation model in macromolecular simulations, *Biopolymers* *56*, 275-291.
53. Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the generalized Born model suitable for macromolecules, *J. Phys. Chem. B* *104*, 3712-3720.
54. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate calculation of hydration free-energies using macroscopic solvent models, *J. Phys. Chem.* *98*, 1978-1988.
55. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules, *Comput. Phys. Commun.* *91*, 1-41.
56. Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011) Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery, *Curr Comput Aided Drug Des.* *7*, 146-157.
57. Morris, G. M., and Lim-Wilby, M. (2008) Molecular docking, *Methods Mol. Biol.* *443*, 365-382.
58. Durrant, J., and McCammon, J. A. (2011) Molecular dynamics simulations and drug discovery, *BMC Biol.* *9*, 71.
59. Blake, L., and Soliman, M. S. (2014) Identification of irreversible protein splicing inhibitors as potential anti-TB drugs: insight from hybrid non-covalent/covalent docking virtual screening and molecular dynamics simulations, *Med Chem Res* *23*, 2312-2323.
60. Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E. M., Sayed, Y., Ibrahim, M. A. A., Naicker, P., and Soliman, M. E. S. (2013) Comparison of the molecular dynamics and calculated binding free energies for nine FDA-approved HIV-1 PR drugs against subtype B and C-SA HIV PR, *Chem. Biol. Drug. Des.* *81*, 208-218.
61. Deng, Y., and Roux, B. (2009) Computations of Standard Binding Free Energies with Molecular Dynamics Simulations, *J. Phys. Chem. B* *113*, 2234-2246.
62. Wang, J., Deng, Y., and Roux, B. (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials, *Biophys. J.* *91*, 2798-2814.
63. Feig, M., and Brooks, C. L. (2002) Evaluating CASP4 predictions with physical energy functions, *Proteins* *49*, 232-245.
64. Ylilauri, M., and Pentikäinen, O. T. (2013) MMGBSA As a Tool To Understand the Binding Affinities of Filamin–Peptide Interactions, *J. Chem. Inf. Model.* *53*, 2626-2633.
65. Soliman, M. E. S., Pernia, J. J. R., Greig, I. R., and Williams, I. H. (2009) Mechanism of glycoside hydrolysis: A comparative QM/MM molecular dynamics analysis for wild type and Y69F mutant retaining xylanases, *Org. Biomol. Chem.* *7*, 5236-5244.

66. Gourmala, C., Luo, Y., Barbault, F., Zhang, Y., Ghalem, S., Maurel, F., and Fan, B. (2007) Elucidation of the LewisX-LewisX carbohydrate interaction with molecular dynamics simulations: A glycosynapse model, *J. Mol. Struct-Theochem* 821, 22-29.
67. Shaikh, S. A., and Jayaram, B. (2007) A swift all-atom energy-based computational protocol to predict DNA-ligand binding affinity and Delta T(m), *J. Med. Chem.* 50, 2240-2244.
68. Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes, *J. Mol. Biol.* 330, 891-913.
69. Hou, T., and Yu, R. (2007) Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: Mechanism for binding and drug resistance, *J. Med. Chem.* 50, 1177-1188.
70. Stoica, I., Sadiq, S. K., and Coveney, P. V. (2008) Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases, *J. Am. Chem. Soc.* 130, 2639-2648.
71. Ode, H., Matsuyama, S., Hata, M., Hoshino, T., Kakizawa, J., and Sugiura, W. (2007) Mechanism of drug resistance due to N88S in CRF01_AE HIV-1 protease, analyzed by molecular dynamics simulations, *J. Med. Chem.* 50, 1768-1777.
72. Perez, M. (2005) Gibbs-Thomson effects in phase transformations, *Scripta Mater.* 52, 709-712.

Chapter 6

Conclusion and future recommendations

This chapter outlines the general conclusion of the study as well as recommendations for future research based on its findings.

6.1. General Conclusion

The research studies reported in this thesis are based on two classes of HIV-1 inhibitors, which include (i) integrase inhibitors and (ii) protease inhibitors: their potential as anticancer agents. To accomplish our objectives, there are two specific aims in this study, which are: (a) To investigate the inhibitory profile of a set of compounds as HIV-1 strand transfer inhibitors and employ a 2D-QSAR approach to predict the biological activities of the ‘test set’ (b) To provide a molecular insight into the binding affinities of HIV-1 protease inhibitors that are found to inhibit Hsp90, the enzyme that initiates the HER2+ breast cancer. To a great extent, this work has accomplished the aims of the studies and results from this work produced the following conclusions:

- 1) In this present study, 40 diketo acid and carboxamide derivatives were analyzed using a QSAR statistical approach. A 2D-QSAR model was developed based on molecular, structural, physicochemical, 2D and 3D properties that were obtained from Chemdraw 10.0 (1) and Discovery studio v3.5. The result suggests that the Radius of gyration, Zagreb index, Wiener index and minimized energy are statistically significant with a correlation coefficient value of 0.820. These descriptors have played an important role in identifying some promising compounds that possess HIV-1 inhibitory properties. The synthesis of the compounds considered in this study was reported in literature (2-4) and were validated using a 2D-QSAR approach in this study. This QSAR model holds good predictive performance with Q^2 values ranging from 0.47 to 0.62 and R^2 of 0.82 that was calculated using the leave one out (LOO) method.
- 2) Some potential HER2+ breast cancer inhibitors have been identified through the process called “repositioning” of existing drugs. This study established the relevance of relative affinities between enzymes and their ligands, as part of the step-by-step process of drug design. The selection of the reference protein (PDB codes 2CG9,

3PRY and 3HJC) crystal structures were used as templates for the construction of a human Hsp90 β homologue. The Ramachandran plot showed the validity of the human Hsp90 homologue where 98% of all residues, including the active site residues, were in the favored region and 99.8% were in the allowed region. This plot showed a total of 20 outliers. The binding pocket of the homologue in complex with HIV-1 protease inhibitors at both the NTD and CTD was investigated. Some active residues have been identified in both terminals in previous studies (5) and this form the basis of investigating and validating these theories by computational methods.

- 3) The convergence of post-dynamic simulation was validated by RMSD, RMSF and potential energy plots. The residue based RMSF shows that the amino acid residues in the region of 216-287 show higher fluctuation in both the NTD and CTD compared to the other amino acid residues which all exhibit lower fluctuations. It was also observed that the amino acids in the mentioned region demonstrated no significant effect on how the human Hsp90 homologue interacted with the HIV-PIs at both terminals.
- 4) The result also shows that the system was well equilibrated wherein the RMSDs of all the nine complexes at the NTD did not exceed 2Å and the system stabilized after 1000 ps MD simulation. On the other hand, the CTD domain shows that all the nine complexes did not exceed 2Å and the system stabilized after 1800 ps MD simulation. In addition, the fluctuations of potential energies at the NTD were <2000 kcal/mol for 5 ns of MD simulation and CTD show that the fluctuations of the potential energy to be \leq 8000 kcal/mol. Overall, it is observed that equilibration at NTD and CTD was achieved, although they happen at different time frame.
- 5) From the above investigation of the human Hsp90 homologue CTD and NTD binding to HIV-1 protease inhibitors, respectively, it can be concluded that these HIV-1 PIs bind more at the NTD than the CTD. The active site residues responsible for binding at the NTD include; Leu43, Asn46, Lys53, Ile91, Asp97, Met93, Asn101, Ser108, Gly109, Phe133 and Thr179. Enough computational evidence has shown that the NTD active residues show better binding to these nine PIs, thereby inhibiting cancer growth. However, the active site residues responsible for binding at the CTD include Gln523, Val534, Ser535, Lys538, Thr595, Tyr596, Gly597, Trp598 and Met602. It is therefore evident from these results that only few PIs such as RTV, LPV, APV, ATV and NFV bind successfully to the CTD at reasonable RMSD range value. Thus, this study suggests further investigation to be carried out on the CTD to identify more

promising drugs that could bind to its active site residues and exhibit excellent pharmacokinetic and pharmacodynamics properties.

It is clear that inhibition of the HIV-1 virus at different developmental stages will have great clinical advantages to HIV-related cases. Although the repositioning of the HIV-1 PIs as anticancer agents have revealed valuable information which could reduce the mortality rate of HER2+ breast cancer patients, the current lack of understanding of the biochemical, microbiological reactions and side effects of these drugs on the patients is a challenge especially in establishing their mechanism of action. However, the results presented herein are valuable indicators of how promising these approaches and methods are for the development of drugs. These research studies will ultimately contribute to better prognosis and increase the survival rates of HIV-1-infected patients as well as HER2+ breast cancer through the discovery, design and repositioning of existing drugs.

6.2. Recommendations and Future Studies

The computational chemistry methods employed in this study provide an efficient and cost effective tool for drug discovery and design. These methods include: Virtual screening, Molecular dynamics (MD) simulation, binding free energy calculation using MM/GBSA method implemented in AMBER 12 and post-dynamic analyses. These methods were employed for the determination of the drug binding modes, inhibitor stability in the enzyme-binding site, determination of energy interactions between the active site residues and ligands and the verification of docking results.

Future studies may include the following:

1. Further validation of these computational results from the implementation of QM (6, 7). This could help explain the molecular interactions with respect to the distribution of motion, determine the type of bond formation between enzymes and their inhibitors, thus providing an insight into their binding mode.
2. Further analysis and synthesis of the HIV-1 inhibitors will allow for more inhibitor selectivity. Study of preclinical pharmacokinetics ADME (Absorption, distribution, metabolism and excretion) for analysis of the newly discovered anticancer agents would have considerable benefits for research in this field (8).

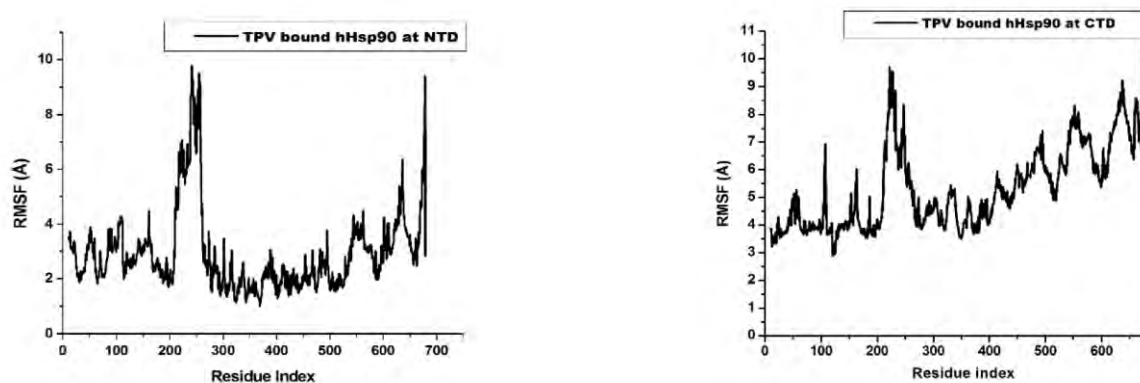
3. Deeper insight into these interactions can be obtained from the implementation of emerging computational methods such as Principal Component Analysis (PCA) (9), Residue Interaction Network (RIN) (10), Substrate Envelope Analysis (SEA) (11) and coarse-grained molecular dynamics (12). Application of these methods could elaborate the enzyme dynamics, drug-enzyme interactions and conformational changes.

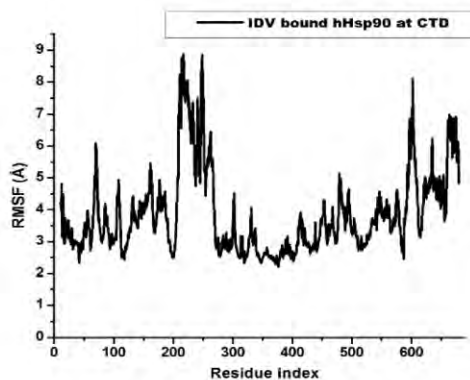
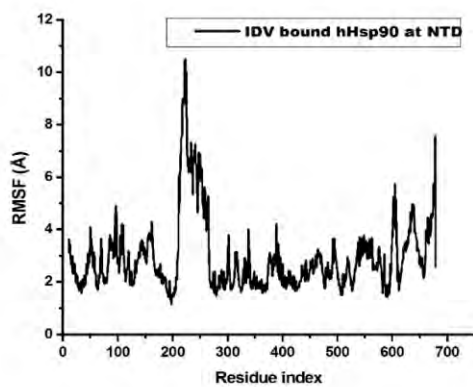
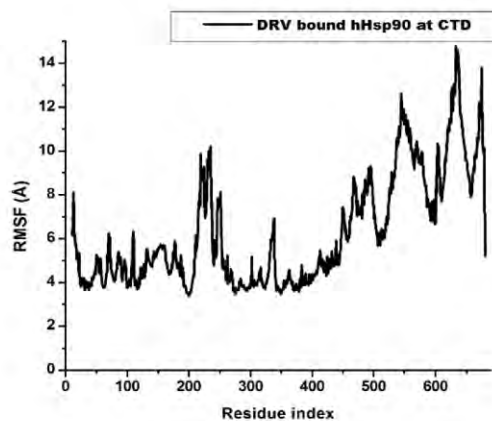
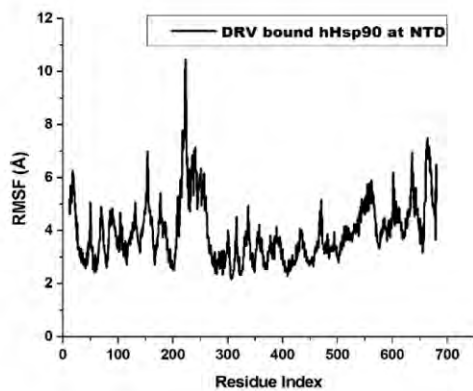
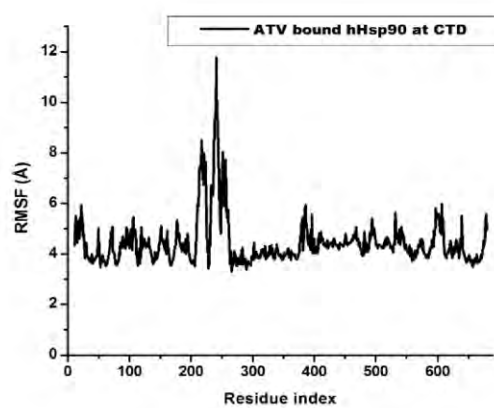
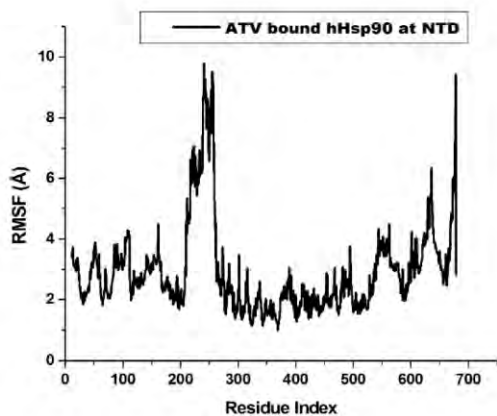
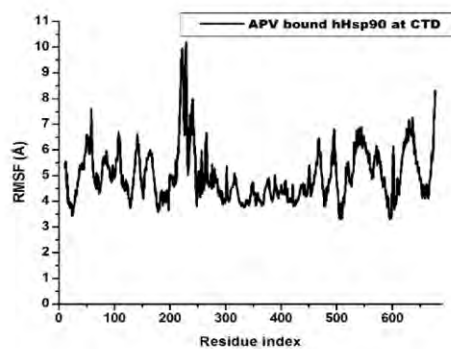
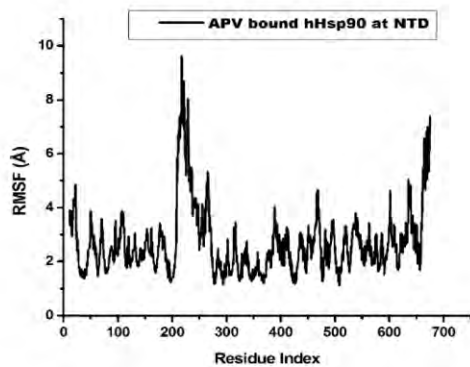
References:

1. Chemdraw. (2010) http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/, 25 Nov. 2013.
2. Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Paz, O. G., Hazuda, D. J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M. V., and Rowley, M. (2008) Discovery of Raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection, *J. Med. Chem.* *51*, 5843-5855.
3. Wai, J. S., Egbertson, M. S., Payne, L. S., Fisher, T. E., Embrey, M. W., Tran, L. O., Melamed, J. Y., Langford, H. M., Guare, J. P., Zhuang, L. G., Grey, V. E., Vacca, J. P., Holloway, M. K., Naylor-Olsen, A. M., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Schleif, W. A., Gabryelski, L. J., and Young, S. D. (2000) 4-aryl-2,4-dioxobutanoic acid inhibitors of HIV-1 integrase and viral replication in cells, *J. Med. Chem.* *43*, 4923-4926.
4. Wai, J. S., Kim, B., Fisher, T. E., Zhuang, L., Embrey, M. W., Williams, P. D., Staas, D. D., Culberson, C., Lyle, T. A., Vacca, J. P., Hazuda, D. J., Felock, P. J., Schleif, W. A., Gabryelski, L. J., Jin, L., Chen, I. W., Ellis, J. D., Mallai, R., and Young, S. D. (2007) Dihydroxypyridopyrazine-1,6-dione HIV-1 integrase inhibitors, *Bioorg. Med. Chem. Lett.* *17*, 5595-5599.
5. Peterson, L. B. (2012) Investigation of the Hsp90 C-terminal binding site, Novel inhibitors and isoform-dependent client proteins, In https://kuscholarworks.ku.edu/dspace/bitstream/1808/10218/1/Peterson_ku_0099D_12224_DATA_1.pdf, 20 Aug. 2013.
6. Mucs, D., and Bryce, R. A. (2013) The application of quantum mechanics in structure-based drug design, *Expert Opin Drug Discov* *8*, 263-276.
7. Van der Kamp, M. W., and Mulholland, A. J. (2013) Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology, *Biochemistry* *52*, 2708-2728.
8. Eddershaw, P. J., Beresford, A. P., and Bayliss, M. K. (2000) ADME/PK as part of a rational approach to drug discovery, *Drug Discov. Today* *5*, 409-414.
9. Maisuradze, G., Liwo, A., and Scheraga, H. (2009) Principal component analysis for protein folding dynamics, *J. Mol. Biol.* *385*, 312-329.
10. Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011) Analyzing and visualizing residue networks of protein structures, *Trends Biochem. Sci.* *36*, 179-182.
11. Shem, Y., Altaman, M. D., Ali, A., Nalam, M. N., Cao, H., Rana, T. M., A., S. C., and Tidor, B. (2013) Testing the substrate-envelope hypothesis with designed pairs of compounds, *ACS Chem. Biol.* *8*, 2433-2441.
12. Bulacu, M., Goga, N., Zhao, W., Rossi, G., Monticelli, L., Periole, X., Tieleman, D. P., and Marrink, S. J. (2013) Improved angle potentials for coarse-grained molecular dynamics simulations, *J. Chem. Theory Comput.* *9*, 3282-3292.

APPENDICES

Appendix 1. Supplementary material for Chapter 5

Could the FDA-approved anti-HIV PR inhibitors be promising anticancer agents? An answer from molecular dynamics analysesOlayide A. Arodola^a, Mbatha H. Sbongile^a and Mahmoud E. S. Soliman^{a*}**Figure 6. Comparative RMSF plot for PIs at the human Hsp90 homologue NTD and CTD**



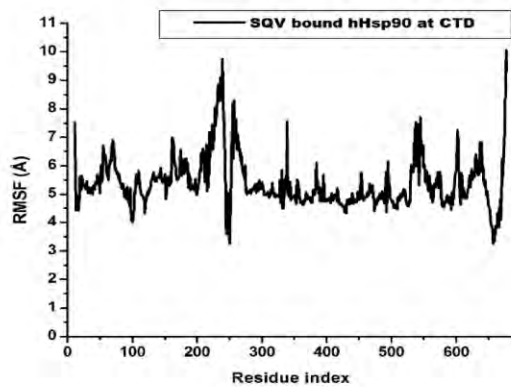
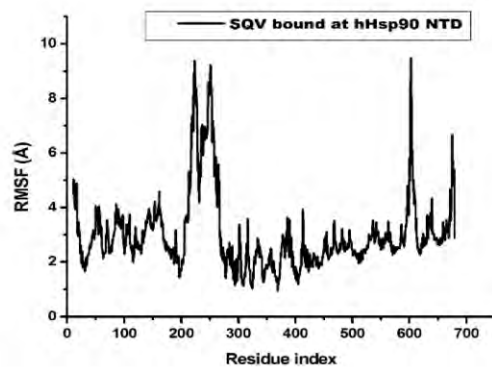
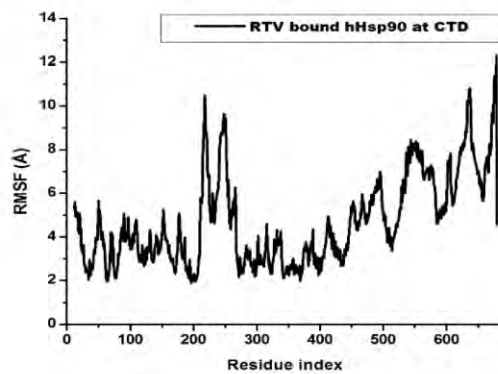
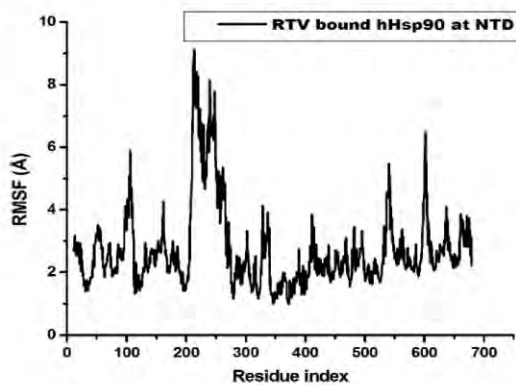
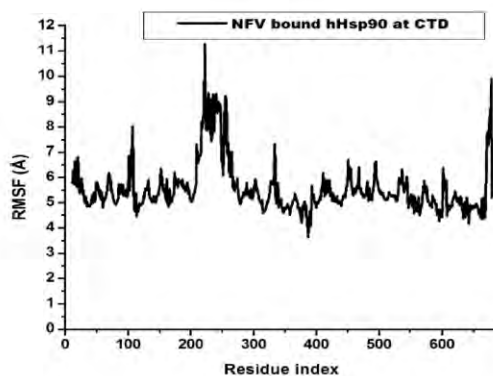
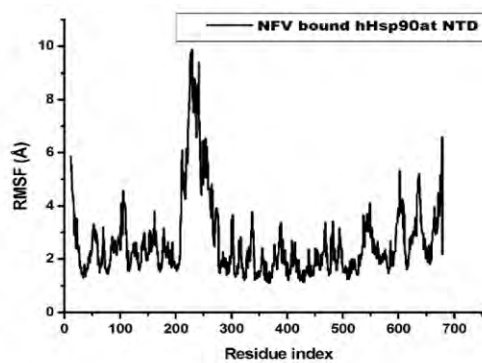
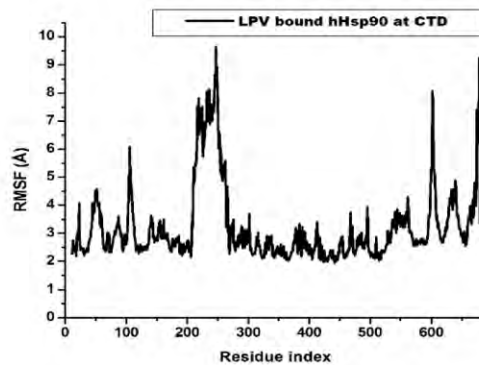
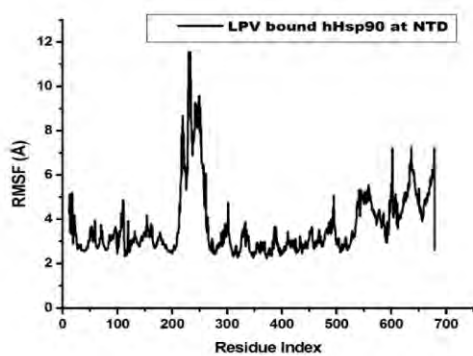
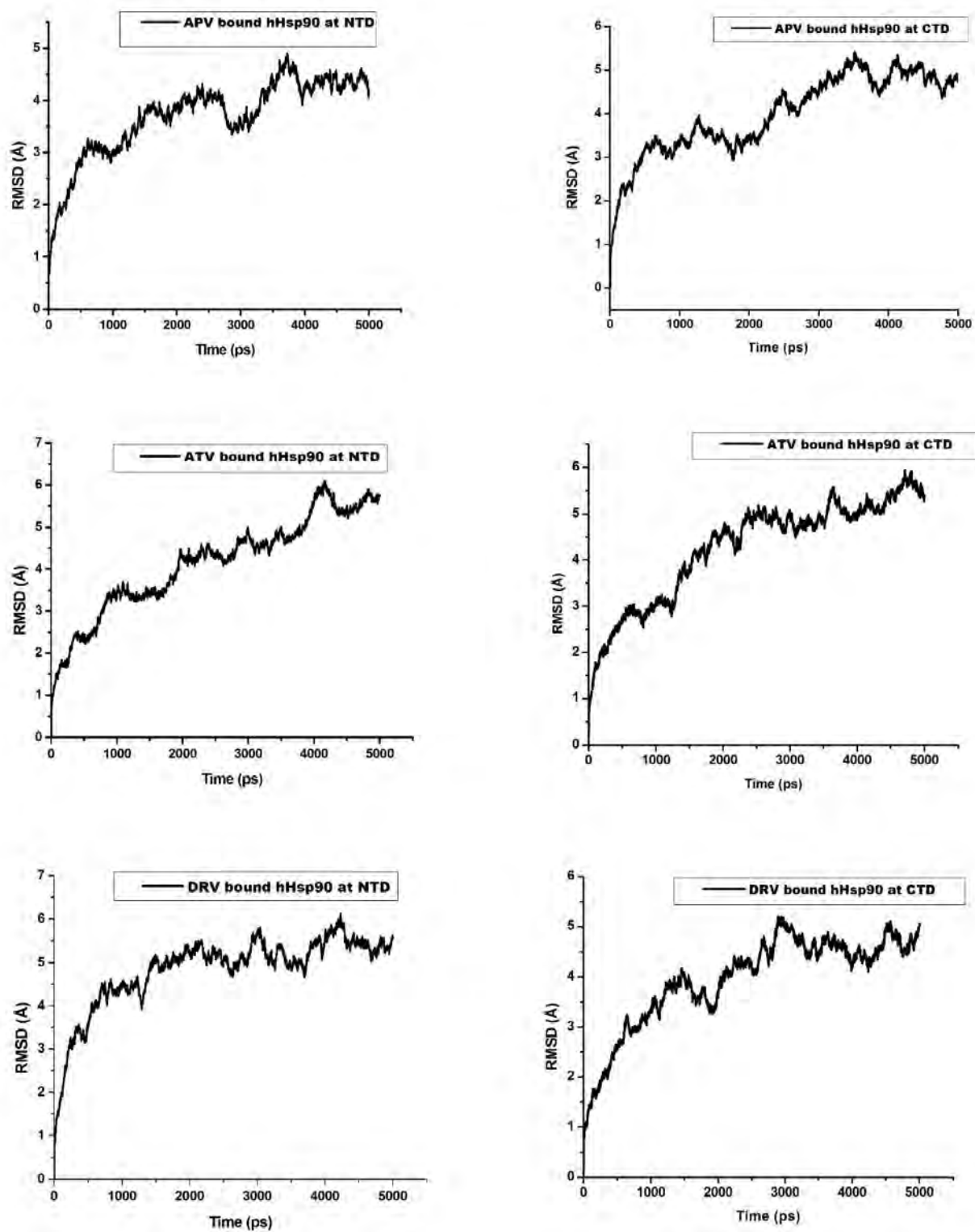
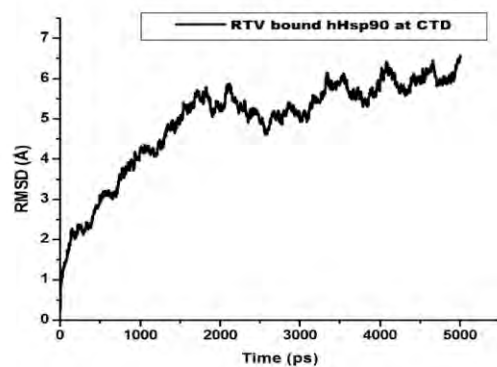
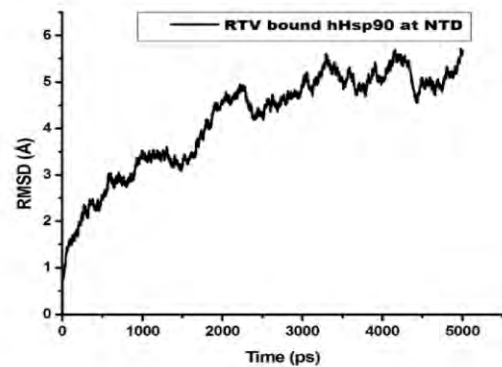
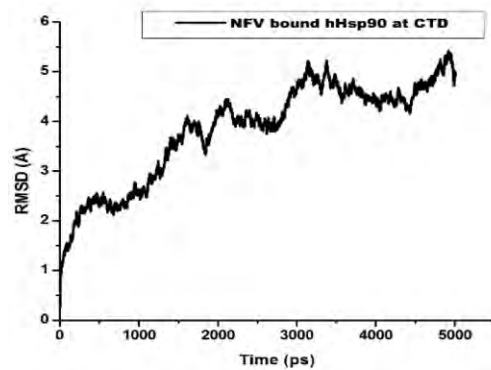
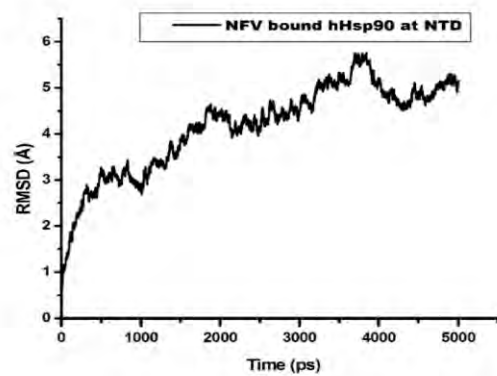
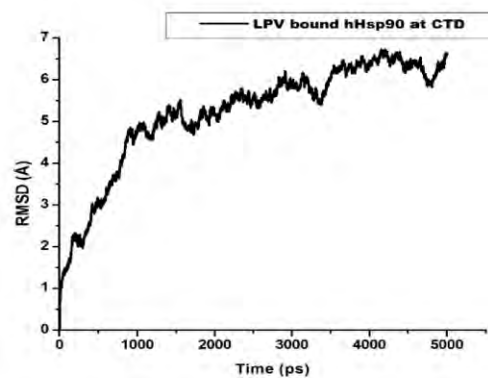
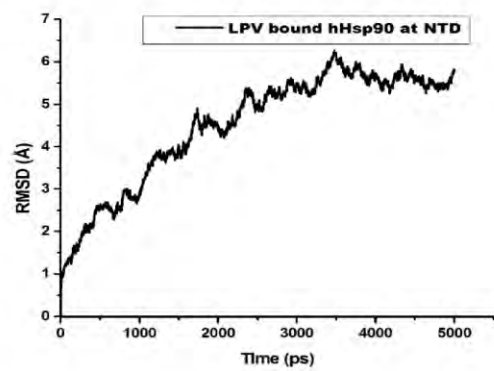
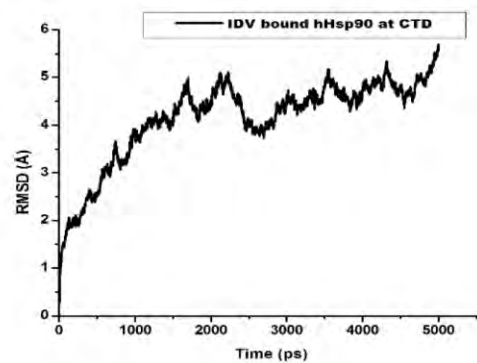
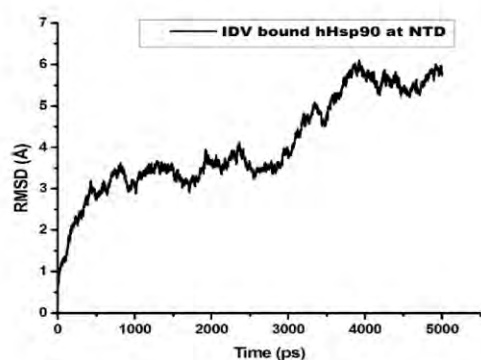


Figure 7a. Comparative RMSD plot for PIs at the human Hsp90 homologue NTD and CTD





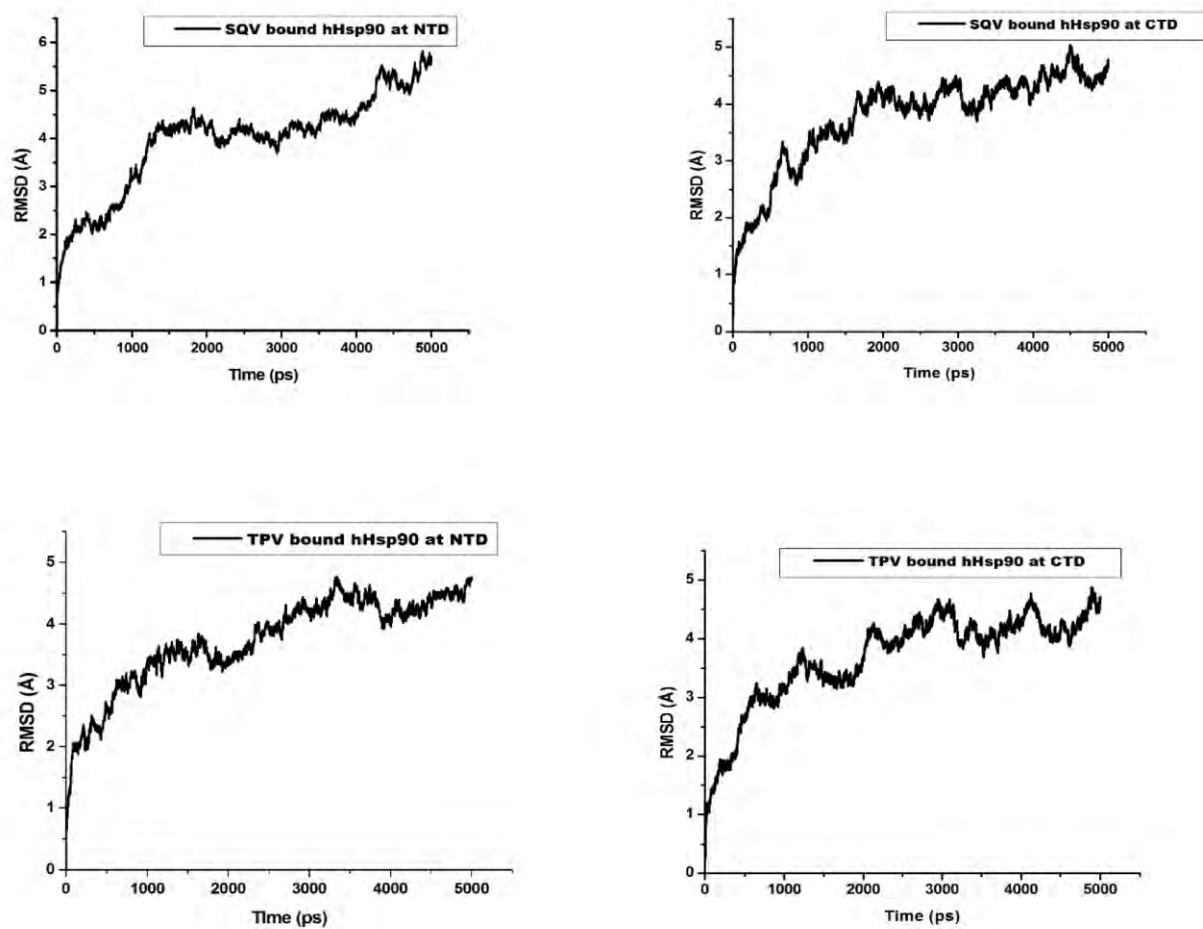
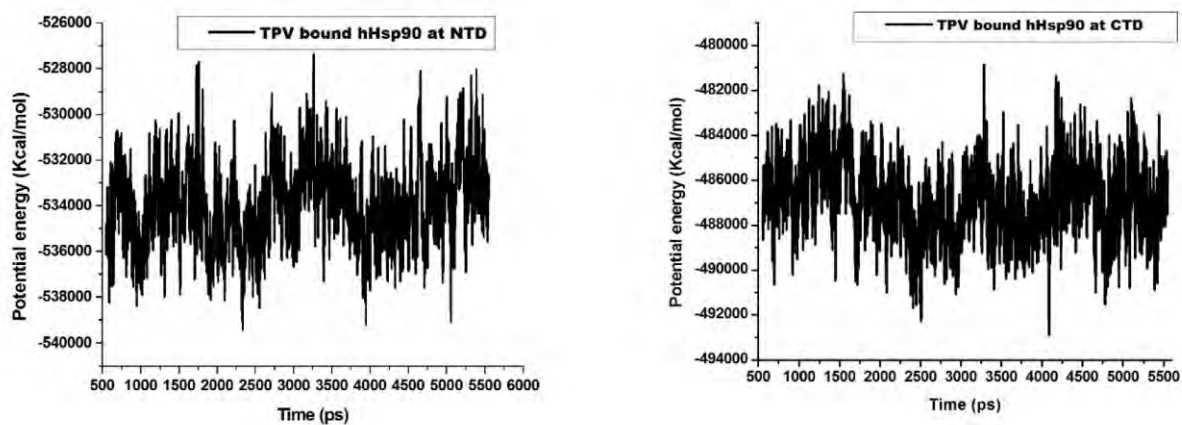
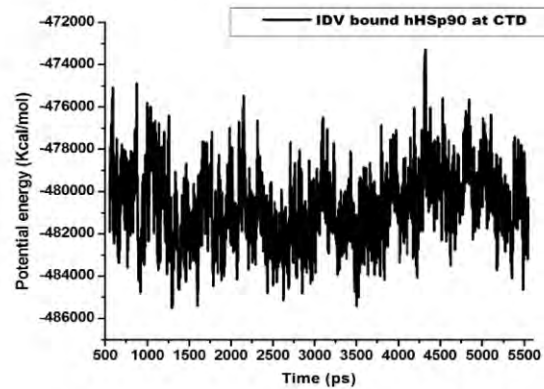
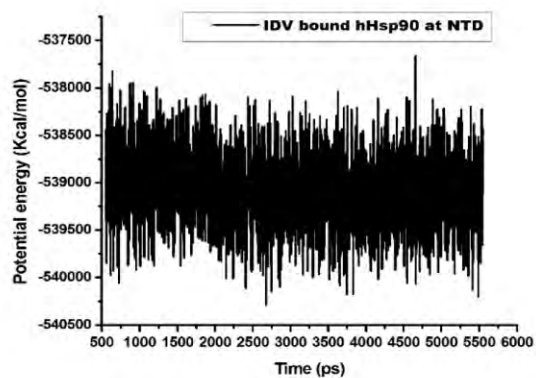
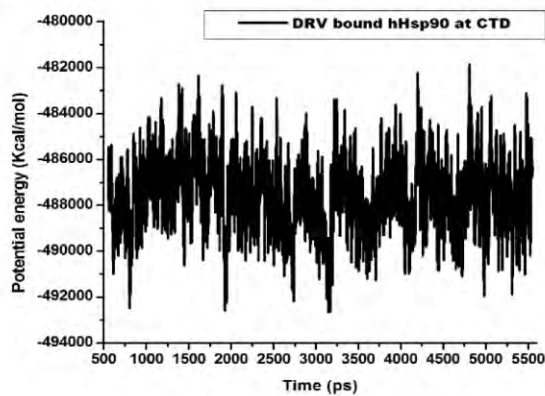
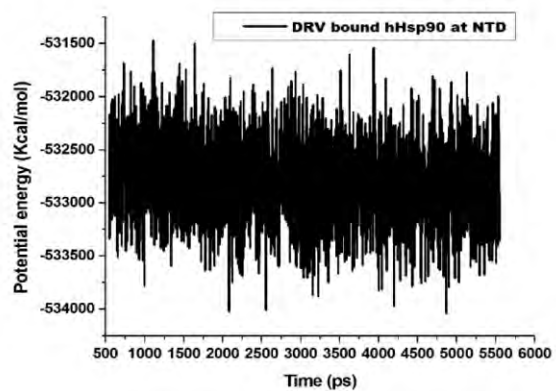
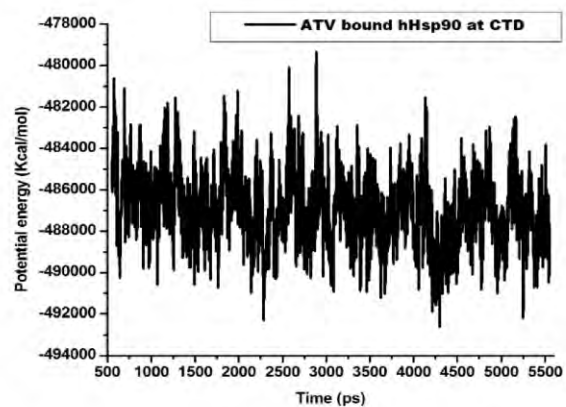
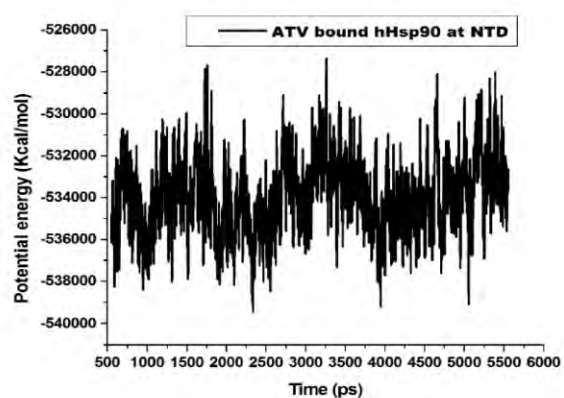
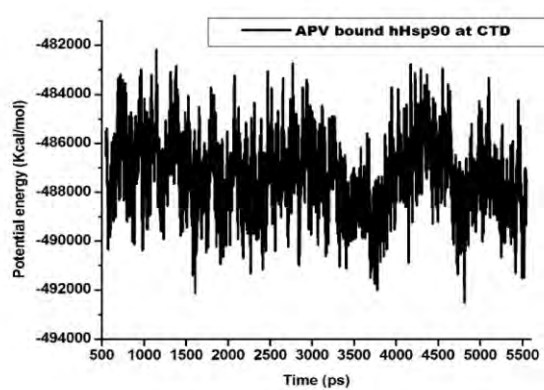
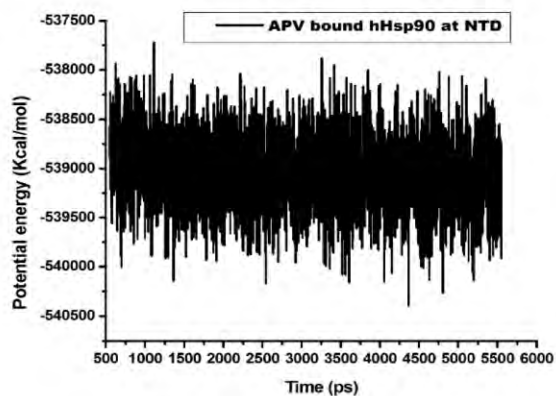
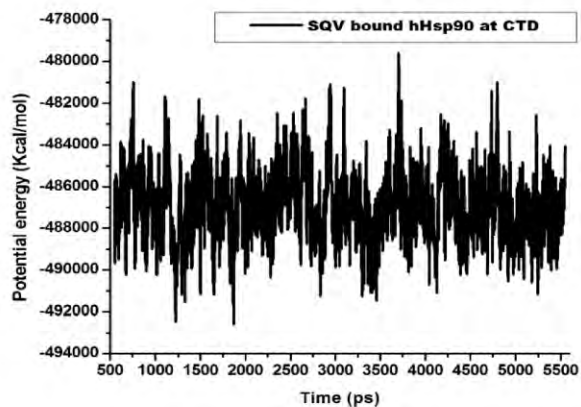
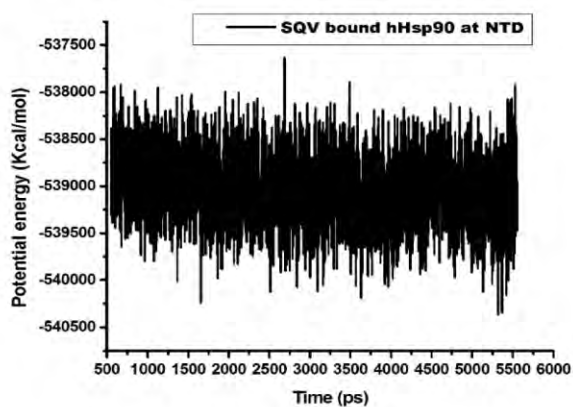
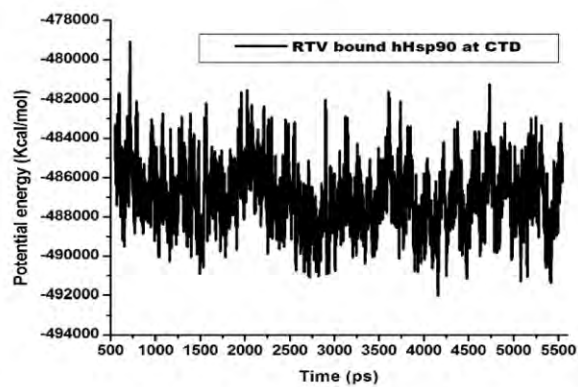
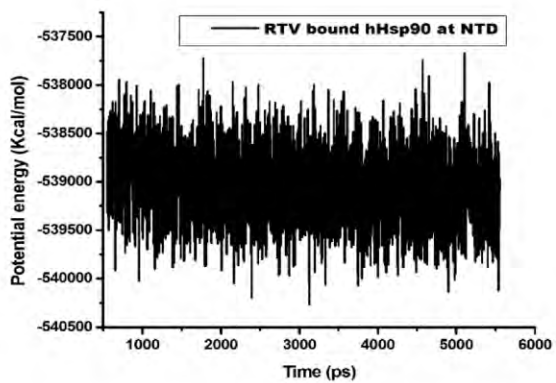
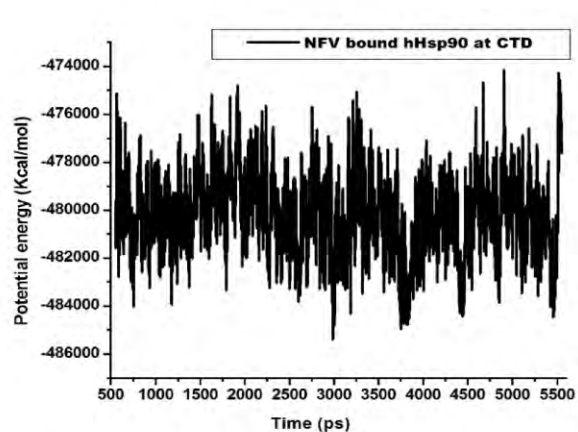
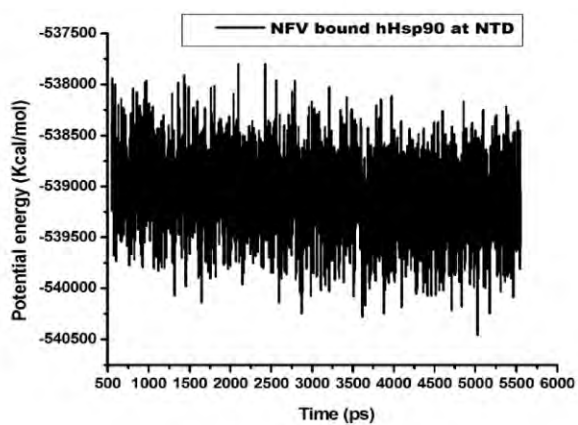
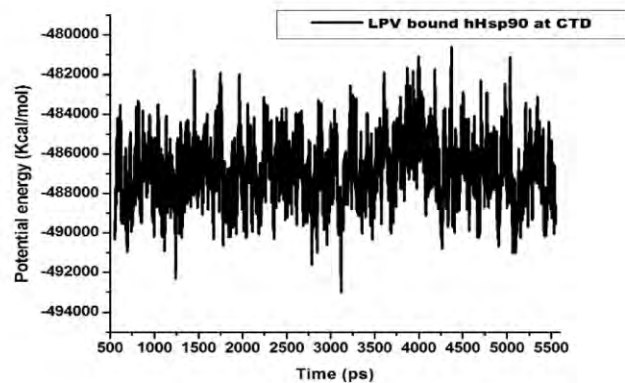
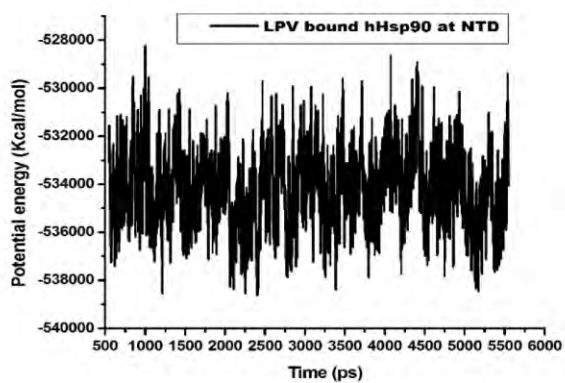


Figure 7b. Comparative potential energy plot for PIs at the human Hsp90 homologue NTD and CTD







Supplementary Figure-S 1. The multiple sequence alignment result from CLUSTAW

CLUSTAL 2.1 multiple sequence alignment

```

3PRY_A|PDBID|CHAIN|SEQUENCE      -----
sp|P08238|HS90B_HUMAN            MPEEVHGGEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNASDA 50
2CG9_B|PDBID|CHAIN|SEQUENCE      -----MASETFEFQAEITQLMSLIINTVYSNKEIFLRELISNASDA 41
3HJC_A|PDBID|CHAIN|SEQUENCE      -----

3PRY_A|PDBID|CHAIN|SEQUENCE      -----
sp|P08238|HS90B_HUMAN            LDKIRYESLTDPSKLDGKELKIDIIPNPQERTLLVDTGIGMTKADLIN 100
2CG9_B|PDBID|CHAIN|SEQUENCE      LDKIRYKSLSDPKQLETEPDLFIRITPKPEQKVLEIRDSGIGMTKAELIN 91
3HJC_A|PDBID|CHAIN|SEQUENCE      -----

3PRY_A|PDBID|CHAIN|SEQUENCE      -----
sp|P08238|HS90B_HUMAN            NLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVVVITKHN 150
2CG9_B|PDBID|CHAIN|SEQUENCE      NLGTIAKSGTKAFMEALSAGADVSMIGQFGVGFYSLFLVADRVQVISKSN 141
3HJC_A|PDBID|CHAIN|SEQUENCE      -----

3PRY_A|PDBID|CHAIN|SEQUENCE      -----
sp|P08238|HS90B_HUMAN            DDEQYAWESSAGGSFTVRADHG-EPIGRGKTVILHLKEDQTEYLEERRVK 199
2CG9_B|PDBID|CHAIN|SEQUENCE      DDEQYIWESNAGGSFTVTLDEVNERIGRGTILRLFLKDDQLEYLEEKRIK 191
3HJC_A|PDBID|CHAIN|SEQUENCE      -----

3PRY_A|PDBID|CHAIN|SEQUENCE      -----
sp|P08238|HS90B_HUMAN            EVVKKHSQFIGYPITLYLEKEREKEISDDEAEEEEKGEEKEDKDDEE--K 247
2CG9_B|PDBID|CHAIN|SEQUENCE      EVIKRHSEFVAYPIQLVVTKEVEKEVPIPEEEKKDDEKKDDEKDDDKK 241
3HJC_A|PDBID|CHAIN|SEQUENCE      -----

3PRY_A|PDBID|CHAIN|SEQUENCE      -----MKTkPIWTRNPDDIT 15
sp|P08238|HS90B_HUMAN            PKIEDVGSDEEDDSGDKKKKTKKIKEYIDQEELNKTkPIWTRNPDDIT 297
2CG9_B|PDBID|CHAIN|SEQUENCE      PKLEEVDEEEE-----KKPKTKKVKEEVQEIEELNKTkPLWTRNPSDIT 285
3HJC_A|PDBID|CHAIN|SEQUENCE      -----MHHHHHSSGRENLyFQGHKPLWTRDPKdVT 31
                                     **:****:*:*

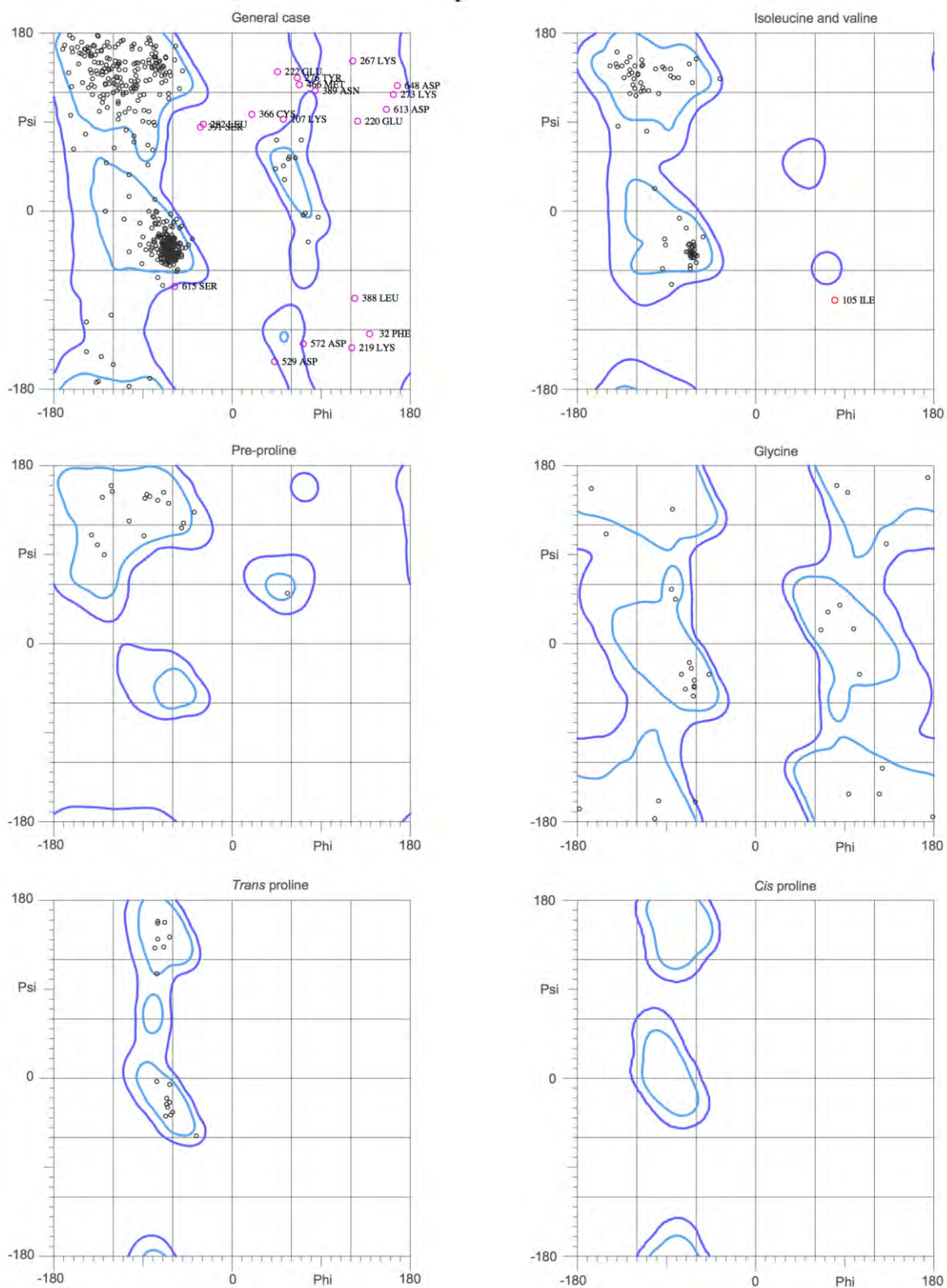
3PRY_A|PDBID|CHAIN|SEQUENCE      QEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFENK 65
sp|P08238|HS90B_HUMAN            QEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFENK 347
2CG9_B|PDBID|CHAIN|SEQUENCE      QEEYNAFYKSLTNDWEDPLYVKHFSVEGQLEFRAILFIPKRAPFDLFESK 335
3HJC_A|PDBID|CHAIN|SEQUENCE      KEEYAIFYKSLTNDWEDPAATKHFSVEGQLEFRSIFVVPKRAPFDLDFEN 81
:***  **:::***** .*****:::~*~:*****:~*~

```

3PRY_A PDBID CHAIN SEQUENCE	KKKNNIKLYVRRVFI	MDSCDELIPEYLN	FIRGVV	SEDLPLNISREMLQQ	115							
sp P08238 HS90B_HUMAN	KKKNNIKLYVRRVFI	MDSCDELIPEYLN	FIRGVV	SEDLPLNISREMLQQ	397							
2CG9_B PDBID CHAIN SEQUENCE	KKKNNIKLYVRRVFI	DEAEDLIPEWLS	FPVKG	VVSEDLPLNLSREMLQQ	385							
3HJC_A PDBID CHAIN SEQUENCE	KKRNNIKLYVRRVFI	MDNCELDLCPDWL	GFVKG	VVSEDLPLNISRNLQQ	131							
	** : *****	* . . . : *	* : . * : *	***** : ***	**							
3PRY_A PDBID CHAIN SEQUENCE	SKILKVIRKNIVK	CLELSELAED	KENYK	KFYAF	SKNLKLG	IHDSTN	165					
sp P08238 HS90B_HUMAN	SKILKVIRKNIVK	CLELSELAED	KENYK	KFYAF	SKNLKLG	IHDSTN	447					
2CG9_B PDBID CHAIN SEQUENCE	NKIMKVIRKNIVK	KLIEAFNEI	AEDE	SEQFE	KFYSA	FNKIKL	GVHEDTQN	435				
3HJC_A PDBID CHAIN SEQUENCE	NKILKVIRKNIVK	CLEMFDEVA	ENKEDY	KQFYEQ	FGKNIK	LGIHEDTAN	181					
	. * : *****	: * * : * : *	* : * : *	* : * : *	* : * : *	* : * : *	* : *					
3PRY_A PDBID CHAIN SEQUENCE	RRRLSELLRYHT	SQSGDEMTSL	SEYVSR	MKETQ	KSIYYIT	TGESKEQ	VANS	215				
sp P08238 HS90B_HUMAN	RRRLSELLRYHT	SQSGDEMTSL	SEYVSR	MKETQ	KSIYYIT	TGESKEQ	VANS	497				
2CG9_B PDBID CHAIN SEQUENCE	RAALAKLLRYN	STKSVDEL	SLTDYV	TRMPE	HQKNI	YYITGES	LKAVEKS	485				
3HJC_A PDBID CHAIN SEQUENCE	RKKLMELLRFY	STESGEEM	TTLKDY	VTRM	KAGQK	SIYYIT	GDSKKKLETS	231				
	* * : * : *	: * : *	* : * : *	* : * : *	* : * : *	* : * : *	: : *					
3PRY_A PDBID CHAIN SEQUENCE	AFVERVRRKGF	EVVYMT	EPIDEY	CVQQL	KFDGK	SLVSVT	KEGLELAEN-	264				
sp P08238 HS90B_HUMAN	AFVERVRRKGF	EVVYMT	EPIDEY	CVQQL	KFDGK	SLVSVT	KEGLELPEDE	547				
2CG9_B PDBID CHAIN SEQUENCE	PFLDALKAKNF	EVFLFL	TDPIDE	YAFTQ	LKEF	EKTLV	DITKD-FE	LEETD	534			
3HJC_A PDBID CHAIN SEQUENCE	PFIEQARRRGL	EVLFMT	EPIDEY	VMQV	KDFED	KKFAC	LTK	EGVHFE	281			
	. * : :	: * : * : *	* : * : *	* : * : *	* : * : *	* : * : *	: : *					
3PRY_A PDBID CHAIN SEQUENCE	-----LYFQ-----							268				
sp P08238 HS90B_HUMAN	EEKKMEESK	AKFENL	CKLMKE	ILDKK	VEKVT	ISNRL	VSSPCC	IVTSTY	597			
2CG9_B PDBID CHAIN SEQUENCE	EEKAEREKE	IKEYE	PLTKAL	KEILG	DQVEK	VVVSY	KLLD	DAPAA	IRTGQ	584		
3HJC_A PDBID CHAIN SEQUENCE	EEKQREEEK	AACEK	LCKTM	KKEVL	GDK	VEK	VIVSER	LSTSP	CILVT	331		
	:											
3PRY_A PDBID CHAIN SEQUENCE	-----											
sp P08238 HS90B_HUMAN	WTANMERIM	KAQAL	RDNST	MGYM	AKKH	LEIN	PDH	PIVET	LRQKA-EADK	646		
2CG9_B PDBID CHAIN SEQUENCE	WSANMERIM	KAQAL	RDSS	SSYMS	S	SKKTFE	ISPK	SPIIK	ELKRR	VDEGGA	634	
3HJC_A PDBID CHAIN SEQUENCE	WSAHMEQ	IMRNQAL	RDSS	MAQ	MMS	SKKTM	ELN	PRH	PIIK	ELRRV-GADE	380	
3PRY_A PDBID CHAIN SEQUENCE	-----											
sp P08238 HS90B_HUMAN	NDKAVKDL	VLLFET	ALLSG	FSLED	PQTH	SNRI	YRM	IKLGL	GID	DEVA	696	
2CG9_B PDBID CHAIN SEQUENCE	QDKTVKDL	TKLLY	ETALL	TSGF	SLDE	PTSF	ASR	INRL	ISL	GLN-----	677	
3HJC_A PDBID CHAIN SEQUENCE	NDKAVKDL	VLLF	DTSL	L	TSGF	QLED	PTG	YER	INR	MIKLGLSL	DEEEEE	430
3PRY_A PDBID CHAIN SEQUENCE	-----											
sp P08238 HS90B_HUMAN	AEEPNA	AVPDE	IPPLE	GDE	DAS	RMEE	VD				724	
2CG9_B PDBID CHAIN SEQUENCE	-----											
3HJC_A PDBID CHAIN SEQUENCE	AAEAPVA	ETAPAEV	-----								444	

The Figure above shows the 2D sequence multi-alignment of Hsp90 from *saccharomyces cerevisiae* (PDB Code: 2CG9), which contained the ATP bound in its active site; Hsp90 middle domain from *homo sapiens* (PDB Code: 3PRY), Hsp90 C- terminal domain from *Leishmania major* (PDB Code: 3HJC) and the human Hsp90 obtained from Uniprot (fasta code P08238).

Supplementary Figure-S 2. Ramachandran plot for the human Hsp90 homologue



91.7% (620/676) of all residues were in favored (98%) regions.

97.0% (656/676) of all residues were in allowed (>99.8%) regions.

There were 20 outliers (phi, psi):

32 PHE (140.0, -124.6)

105 ILE (80.7, -90.8)

107 LYS (53.0, 93.3)

219 LYS (121.6, -138.6)

220 GLU (127.9, 92.0)

222 GLU (46.8, 141.4)

267 LYS (122.1, 153.0)

273 LYS (164.0, 118.7)

276 TYR (66.5, 135.4)

282 LEU (-29.3, 88.2)

366 CYS (20.7, 98.3)

388 LEU (124.5, -88.7)

389 ASN (84.2, 123.0)

391 SER (-32.3, 85.1)

466 MET (68.7, 128.0)

529 ASP (43.6, -152.6)

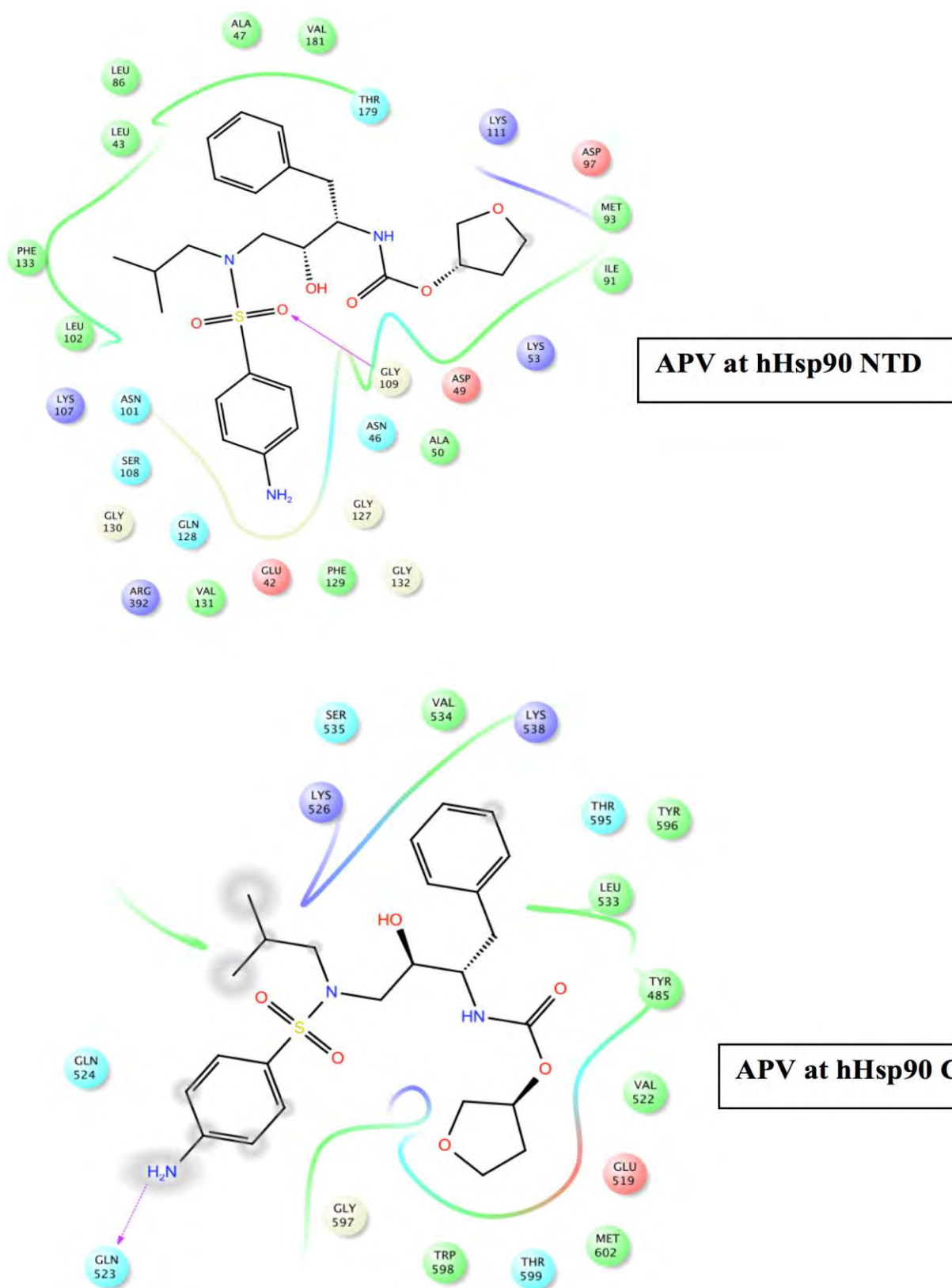
572 ASP (72.8, -134.9)

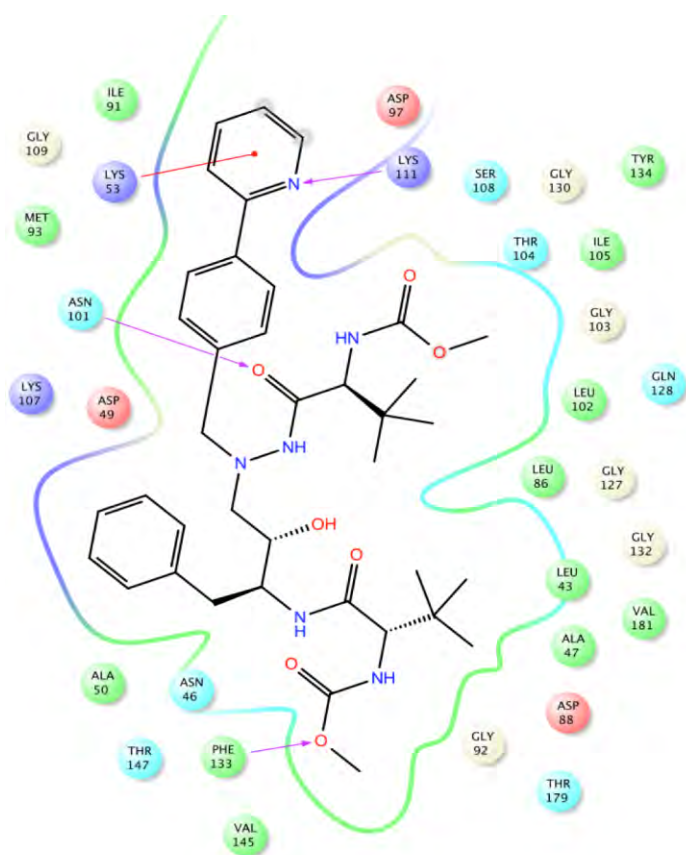
613 ASP (156.6, 103.0)

615 SER (-58.5, -76.8)

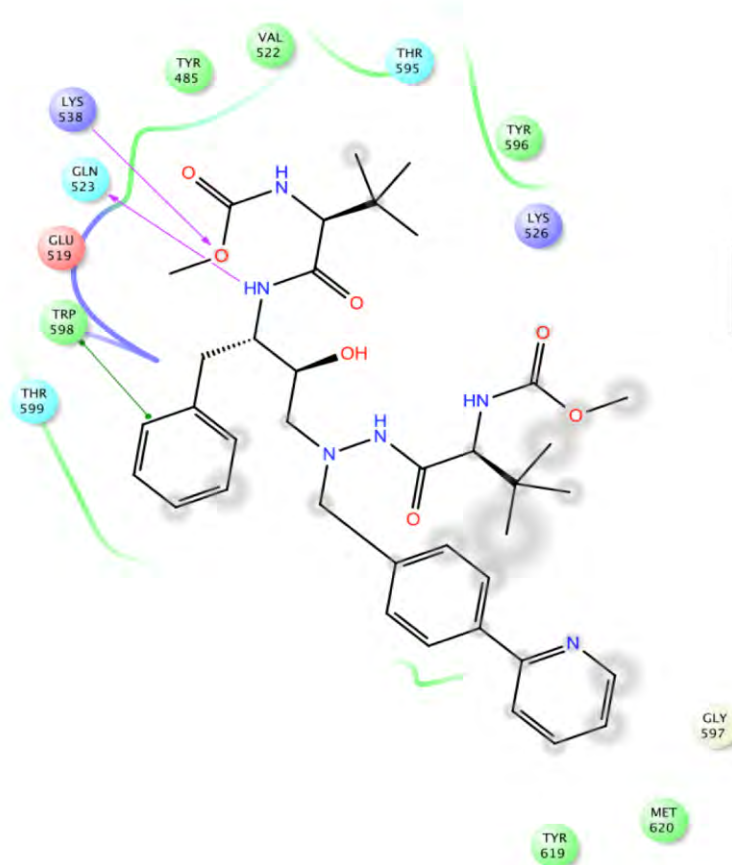
648 ASP (167.7, 127.8)

Supplementary Figure-S 3. Ligand-enzyme interaction at the human Hsp90 homologue NTD and CTD

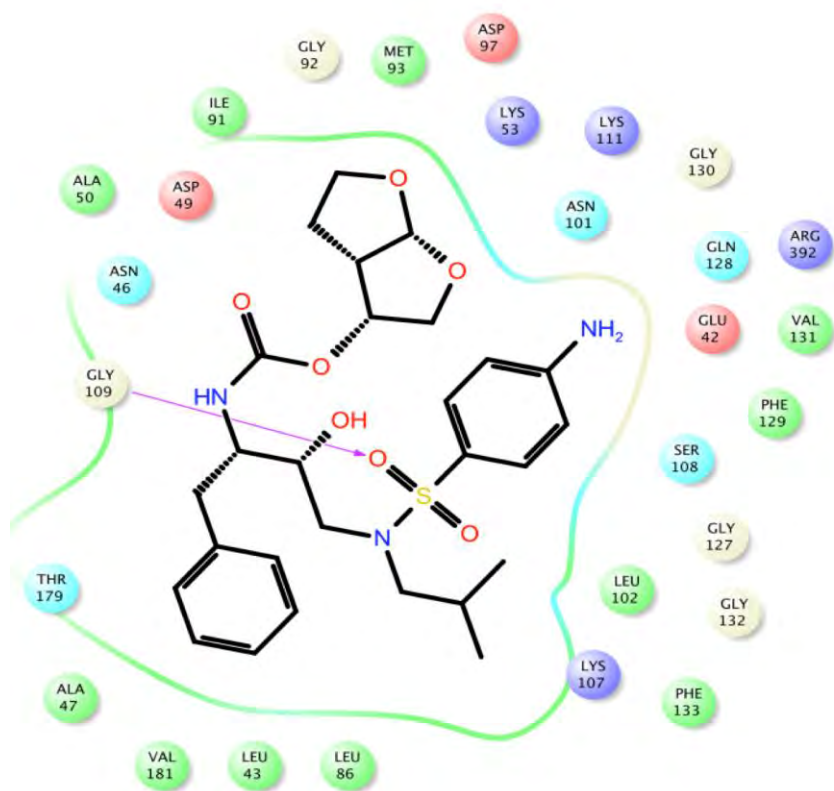




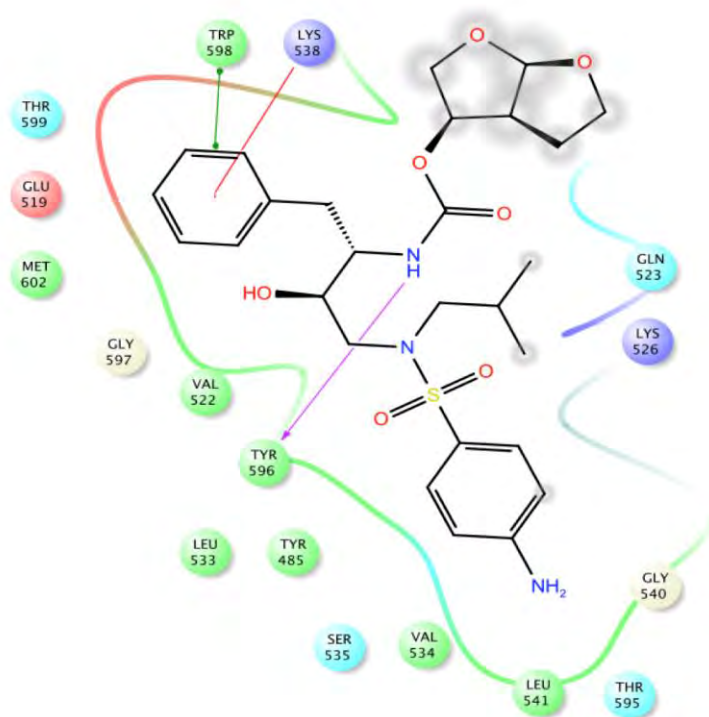
ATV at hHsp90 NTD



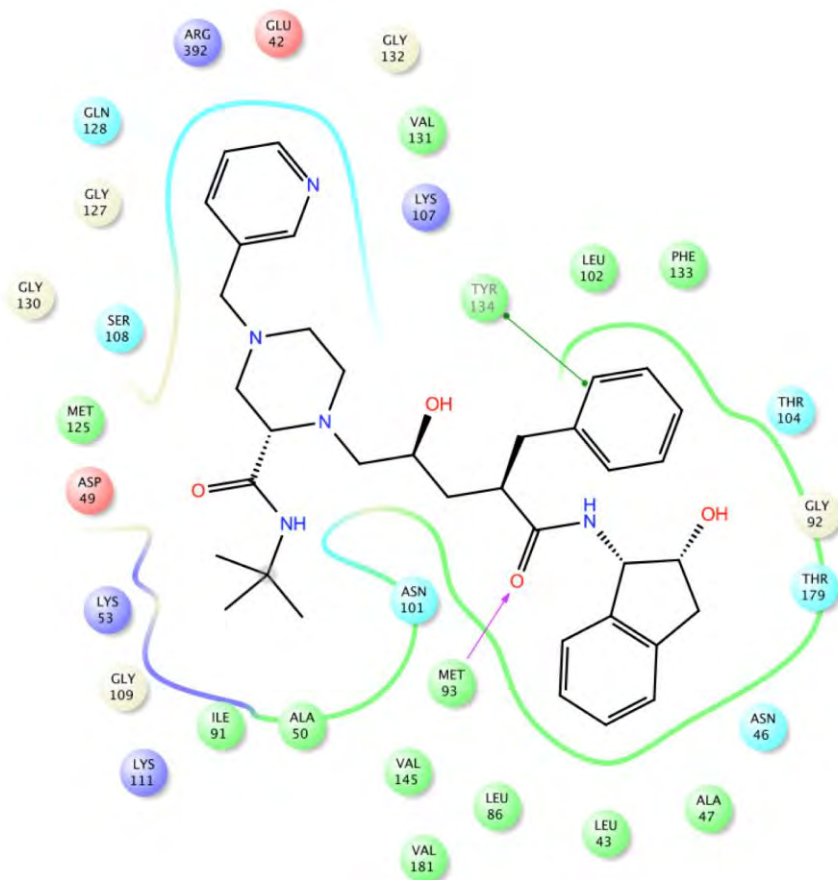
ATV at hHsp90 CTD



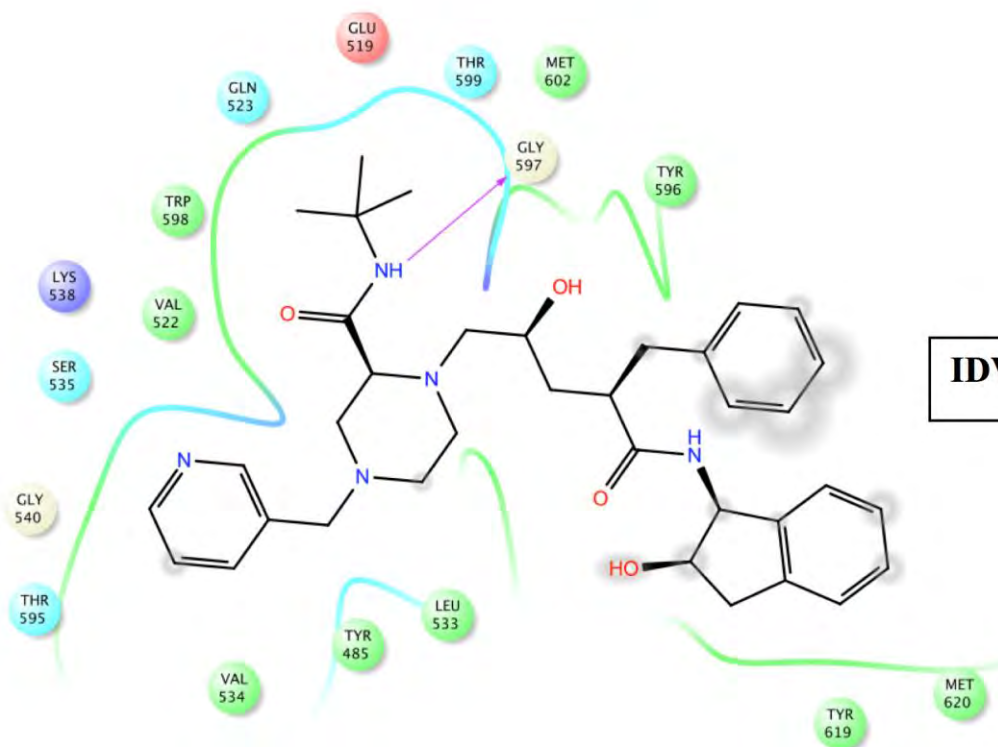
DRV at hHsp90 NTD



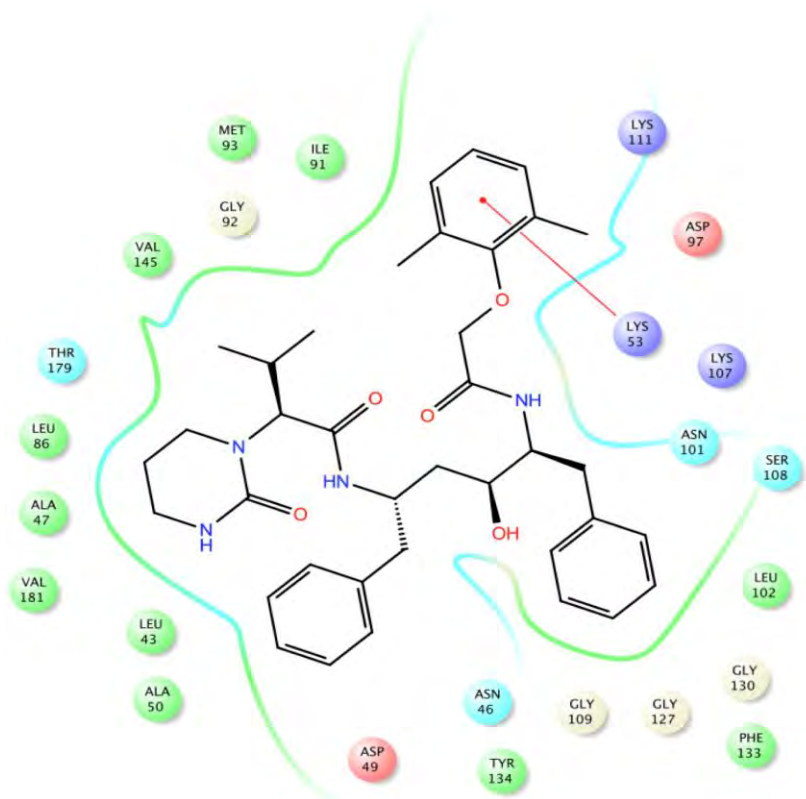
DRV at hHsp90 CTD



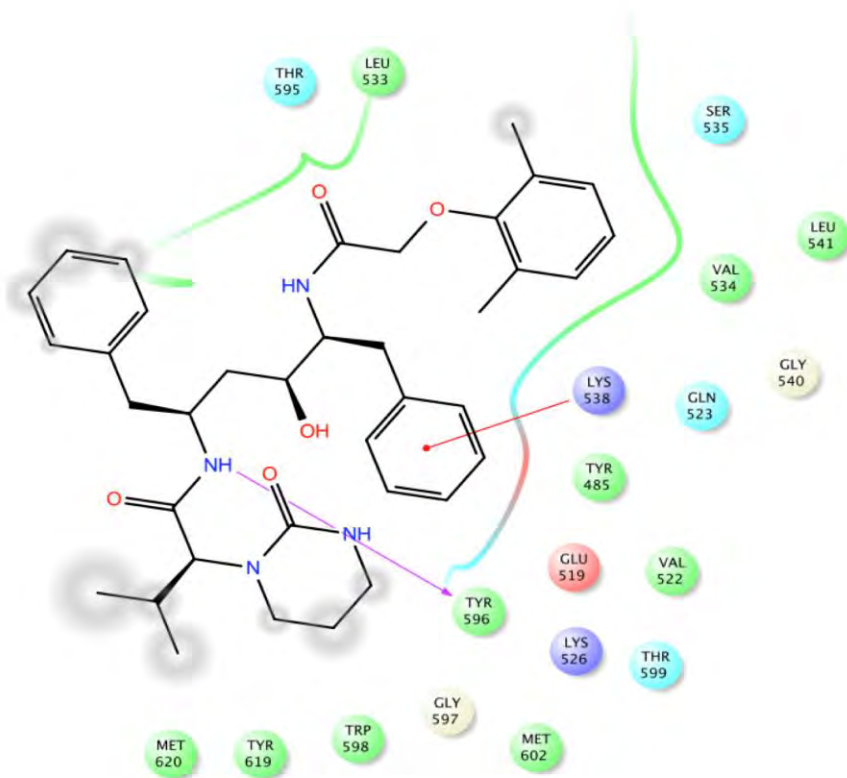
IDV at hHsp90 NTD



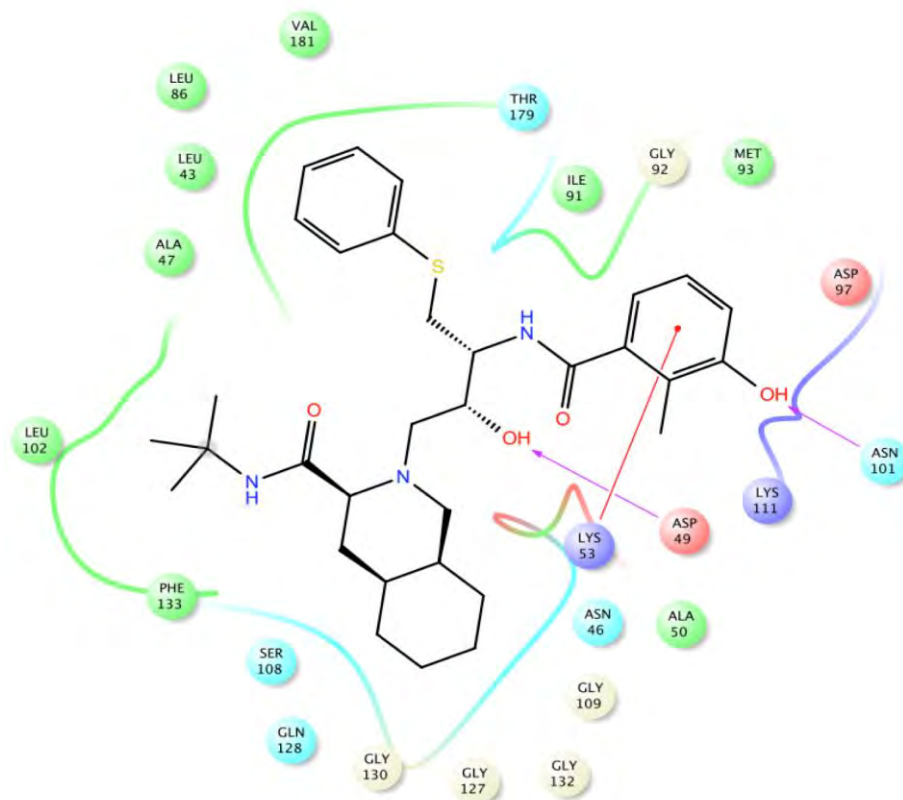
IDV at hHsp90 CTD



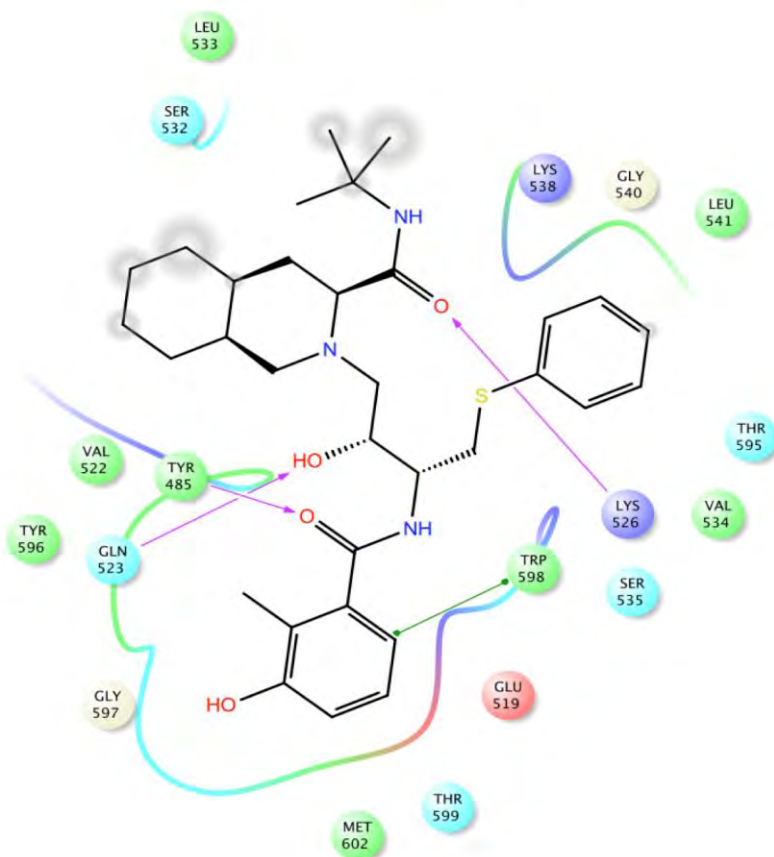
LPV at hHsp90 NTD



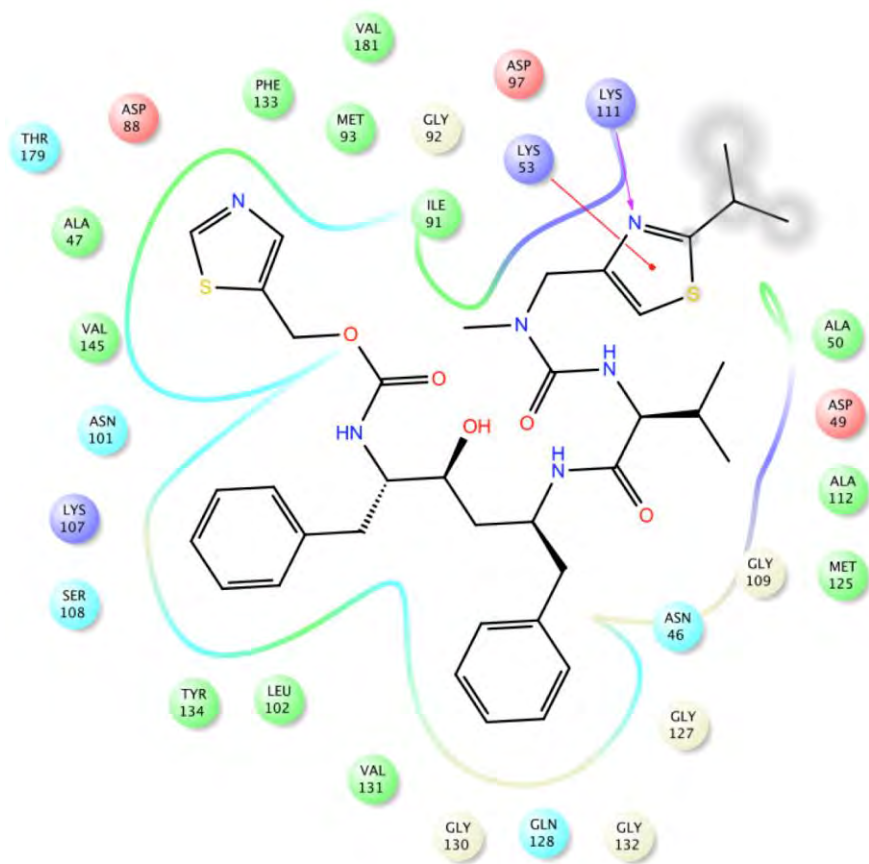
LPV at hHsp90 CTD



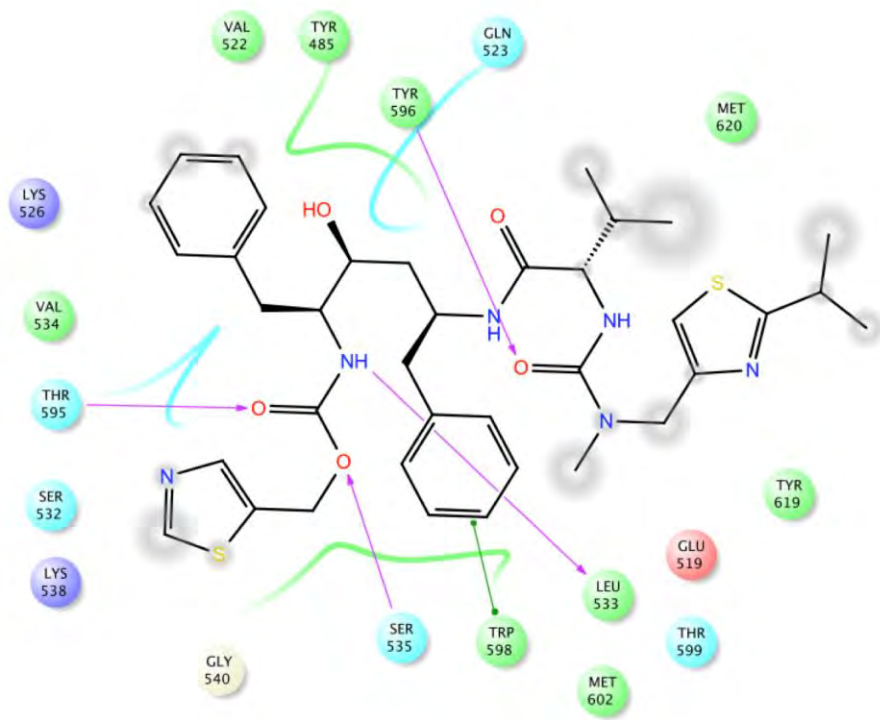
NFV at hHsp90 NTD



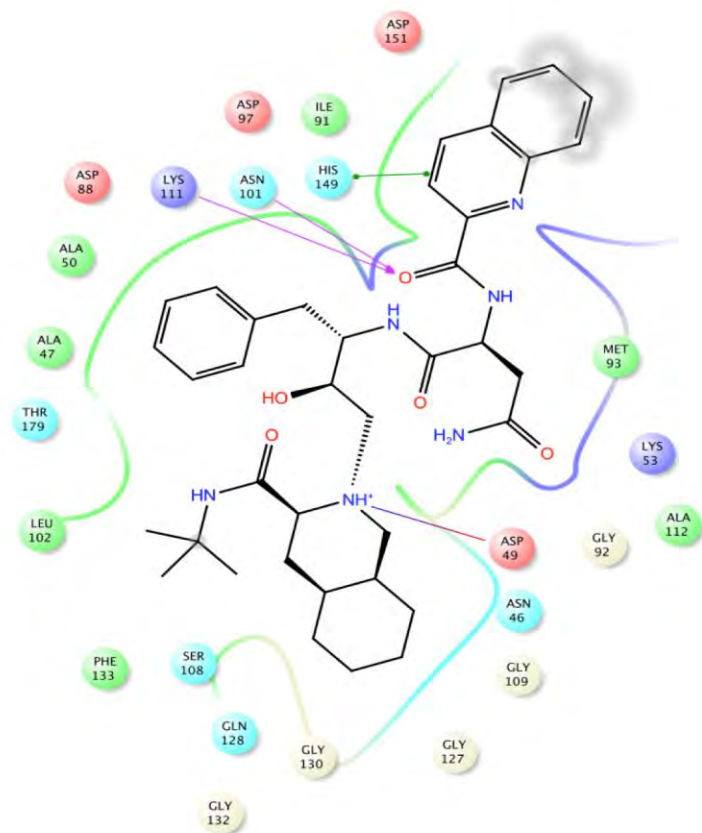
NFV at hHsp90 at CTD



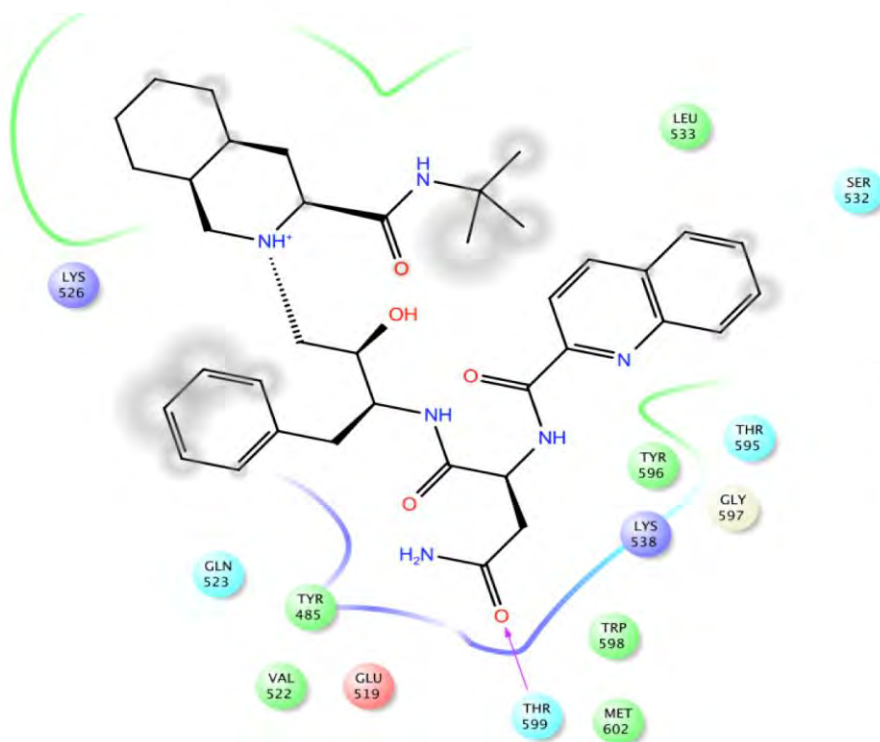
RTV at hHsp90 NTD



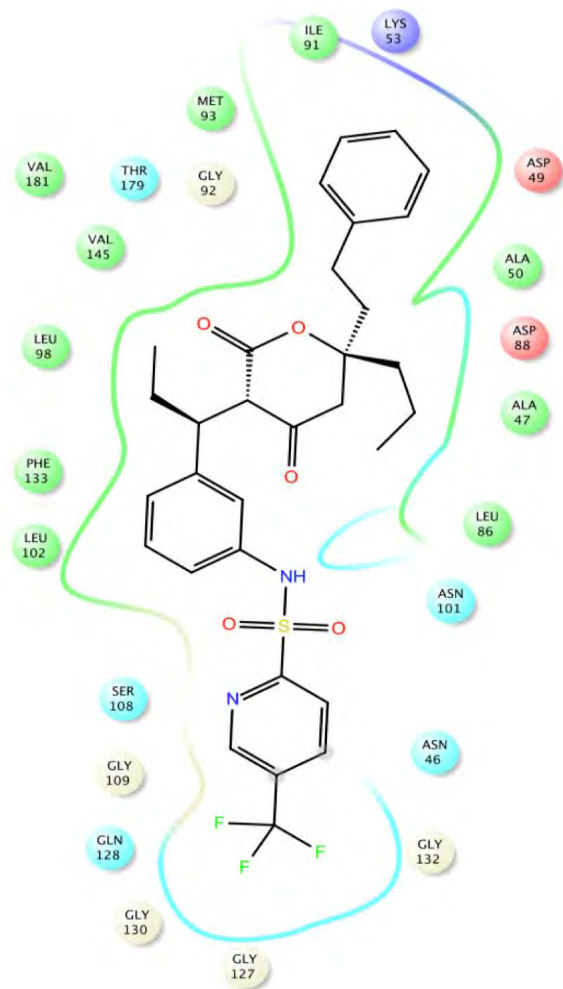
RTV at hHsp90 CTD



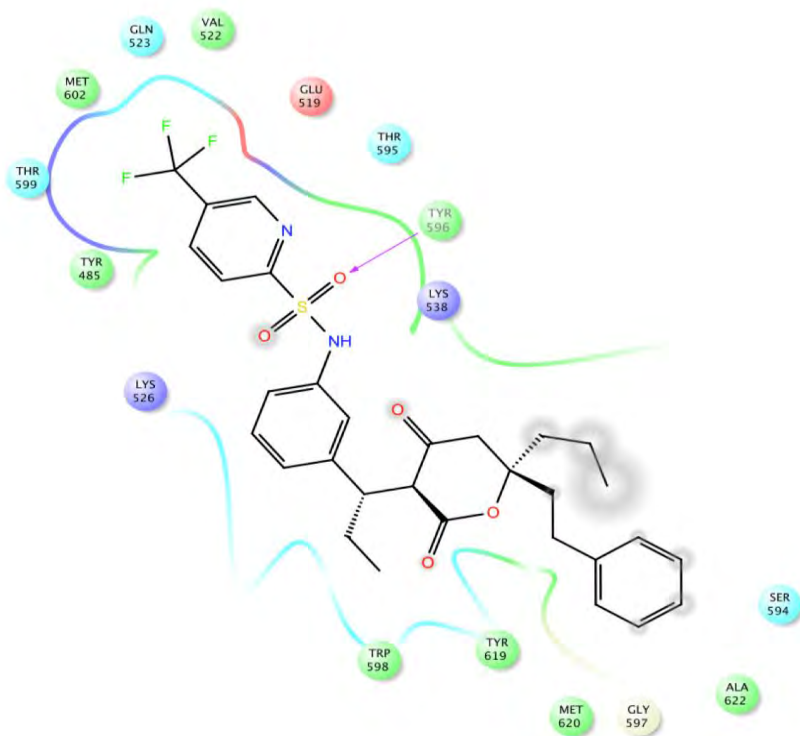
SQV at hHsp90 NTD



SQV at hHsp90 CTD



















TPV at hHsp90 NTD

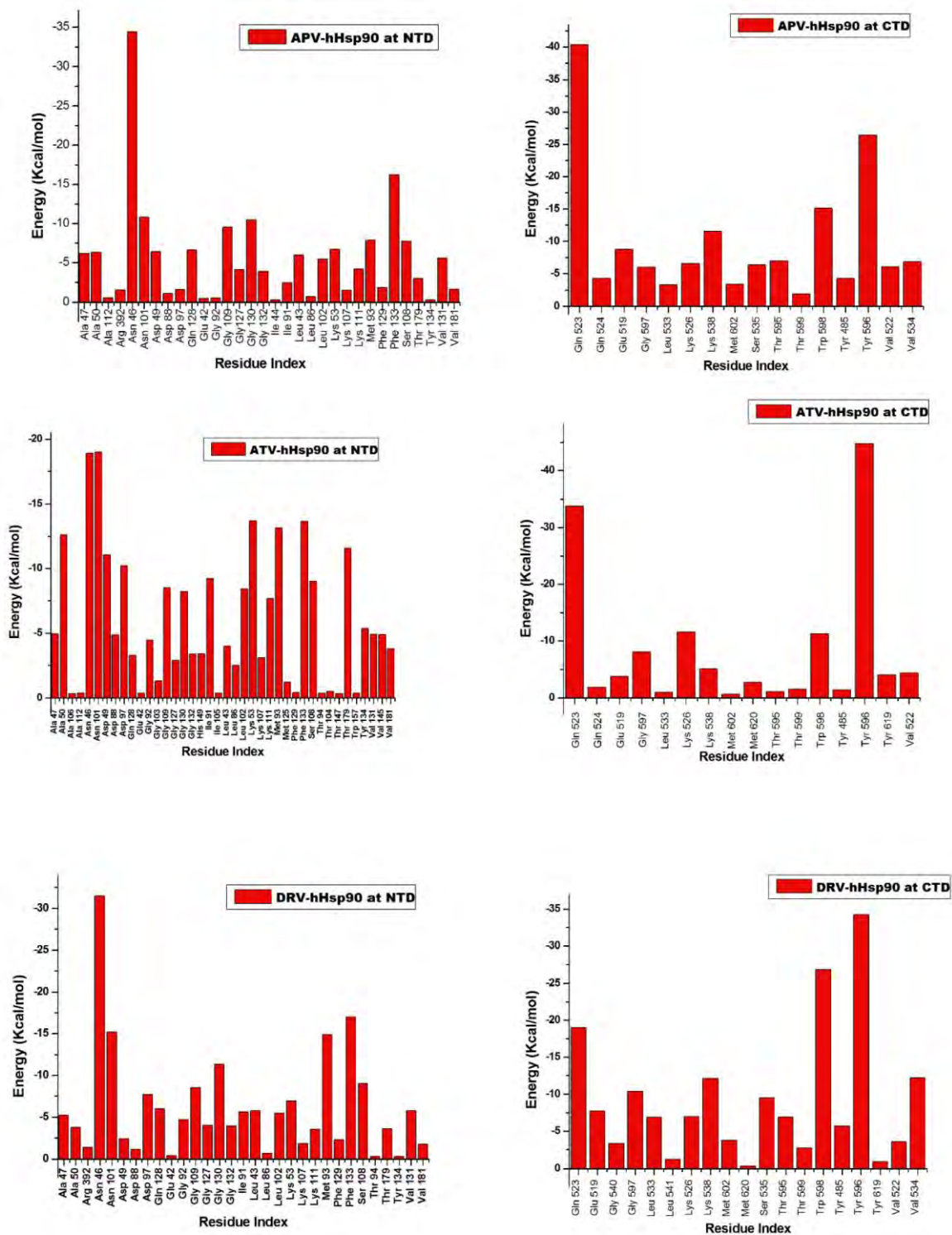


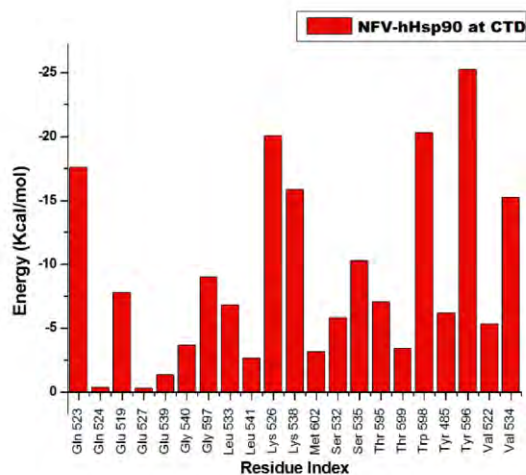
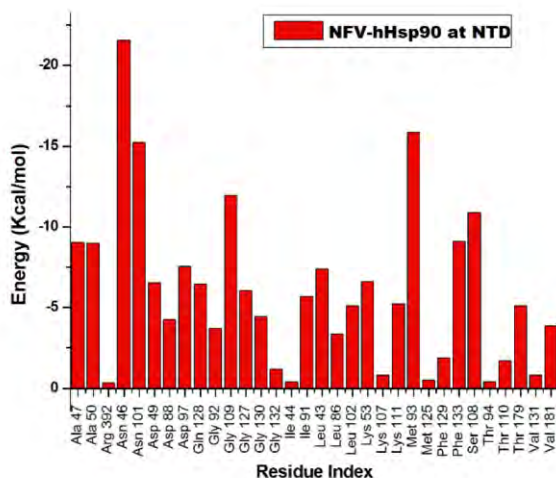
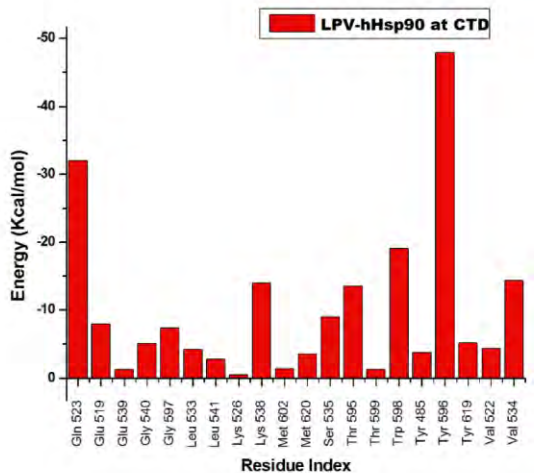
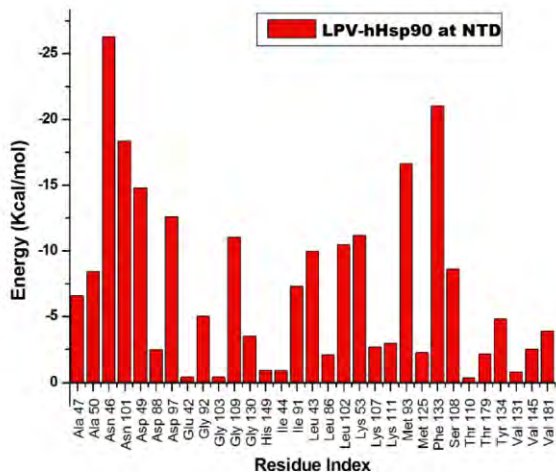
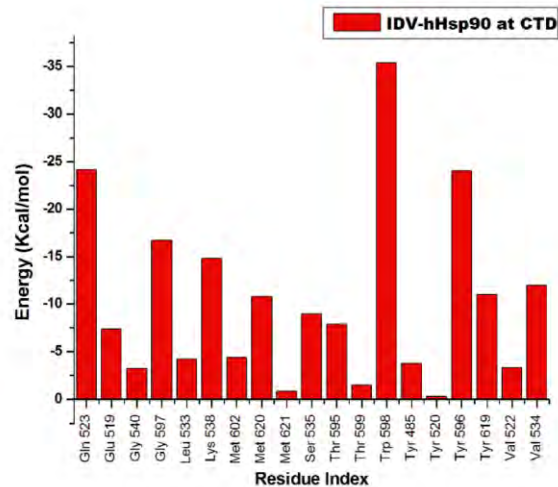
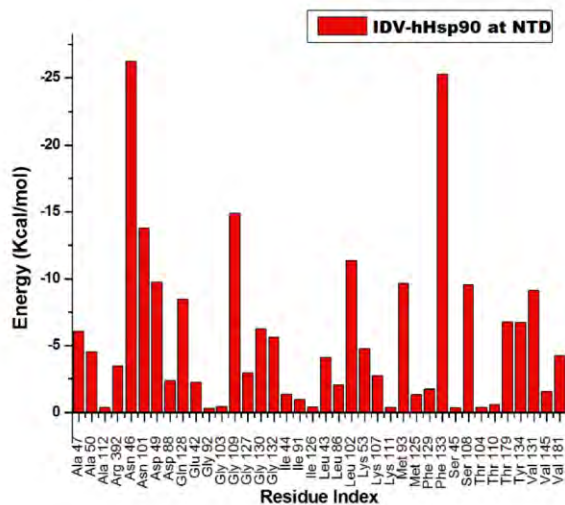
TPV at hHsp90 CTD

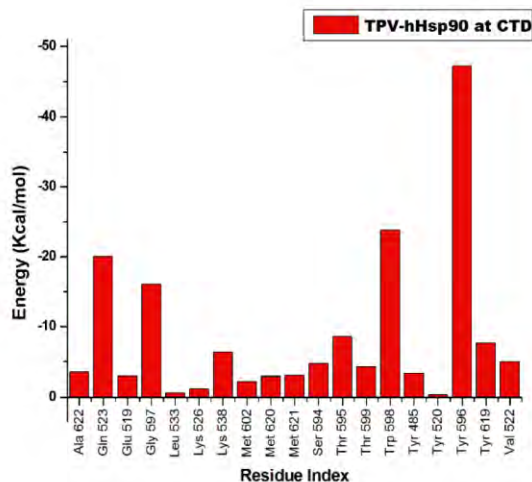
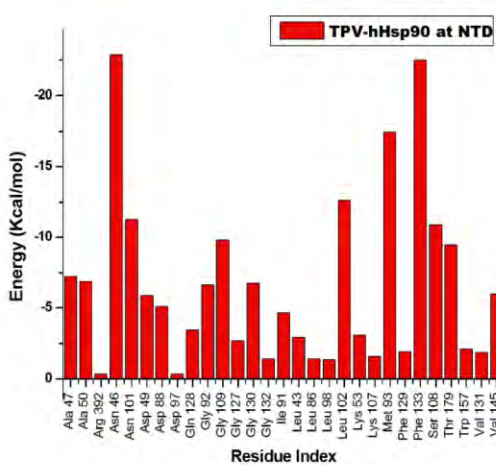
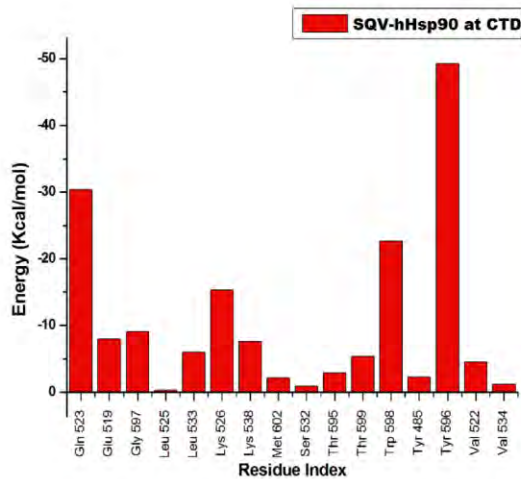
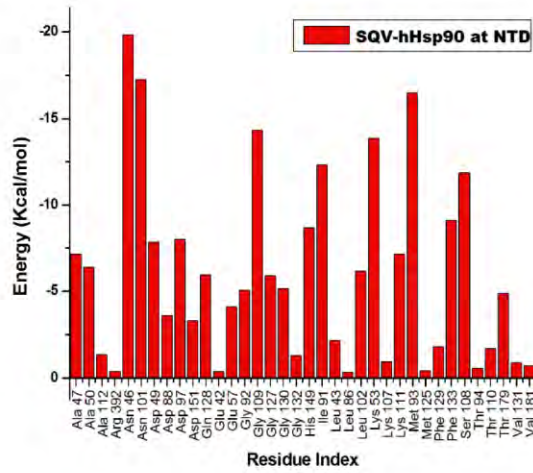
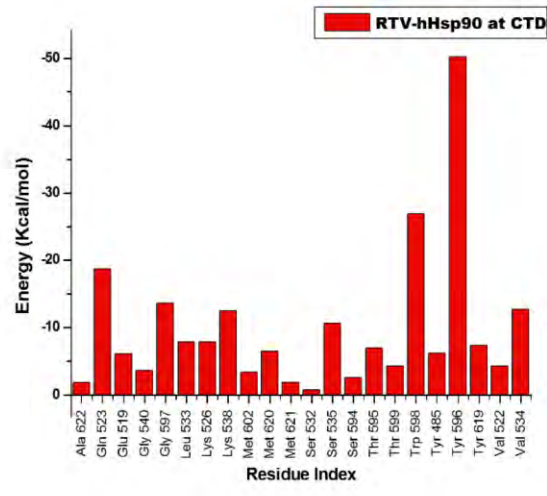
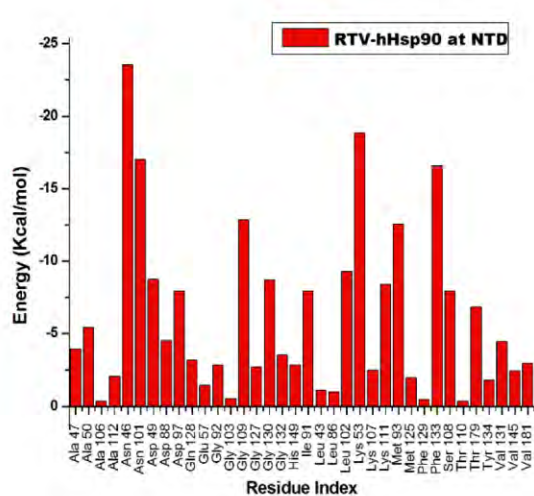
The legend for the ligand-enzyme interaction generated from Ligplot

 Charged (negative)	 Water	 H-bond (side chain)
 Charged (positive)	 Hydration site	 Metal coordination
 Polar	 Displaced hydration site	 Salt Bridge
 Hydrophobic	 π - π stacking	 Solvent exposure
 Glycine	 π -cation	
 Metal	 H-bond (backbone)	

Supplementary Figure-S 4. Comparative per-residue decomposition energy at the human Hsp90 NTD and CTD







Appendix 2. Input files for docking the human Hsp90 homologue complexes at NTD

```
receptor = rec.pdbqt  
exhaustiveness = 8  
center_x = -126.0  
center_y = -33.0  
center_z = 110.0  
size_x = 22  
size_y = 22  
size_z = 20
```

Appendix 3. Input files for docking the human Hsp90 homologue complexes at CTD

```
receptor = rec.pdbqt  
exhaustiveness = 8  
center_x = -81.0  
center_y = -54.0  
center_z = 56.0  
size_x = 16.0  
size_y = 22.0  
size_z = 28.0
```

Appendix 4. PDF version of the published paper