IDENTIFICATION OF ANTI-CANCER AGENTS USING INTEGRATED COMPUTATIONAL TOOLS.

UNIVERSITY OF KWAZULU-NATAL ™

INYUVESI YAKWAZULU-NATALI

SBONGILE HAPPYNESS MBATHA

208526895

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

Supervisor

Prof. Mahmoud Soliman

Durban

2015

IDENTIFICATION OF ANTI-CANCER AGENTS USING INTEGRATED COMPUTATIONAL TOOLS.

2015

Sbongile H. Mbatha
208526895

A thesis submitted to the School of Health Science, University of KwaZulu-Natal, Westville, for the degree of Master of Medical Science (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. Typically these chapters will have been published in internationally recognized, peer-reviewed journals.

This is to certify that the contents of this thesis are the original research work of Ms. Sbongile H. Mbatha.

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

<u>Supervisor:</u>

Signed**:……………………**   Name: **Prof. Mahmoud Soliman**   Date**:……………………**

ABSTRACT

Cancer is a heterogeneous disease that is responsible for various molecular changes and pathological entities that play vital roles in its response to treatment, survival, and growth. Better understanding into the critical pathways and molecular events involved in cancer has enabled the identification of novel targets and development of anti-cancer therapies. In this study, two enzymes that have been shown to be involved in different stages of cancer were used, namely cathepsin B and Hsp90. The main aim of this study was to employ integrated *in-silico* approaches to study inhibitory routes of these enzymes to develop novel anti-cancer agents.

Cathepsin B is the most well studied of the cathepsin family as a potential therapeutic target to treat cancer. To accomplish practical aim 1 of this study, the Michael-acceptor type compounds obtained from a chemical database that irreversibly inhibit cathepsin B were investigated. Validation was carried out using compounds with experimentally determined anti-cathepsin B activity. Four novel compounds exhibited better covalent binding affinity when compared against the experimentally determined prototypes. Molecular dynamics simulations were performed to ensure the stability of the docked complexes and to allow further analysis. Per-residue interaction decomposition analysis provided deeper insight into the interaction themes with active site residues. It was found that polar and hydrophobic interactions had the highest contribution towards drug binding.

Recent experimental studies have documented FDA-approved protease inhibitors involvement in anti-cancer activity, however, there was limited understanding to the mode of inhibition. To accomplish practical aim 2 of this study, the mode of inhibition of protease inhibitors against Hsp90 cysteine protein was investigated. The lack of an X-ray crystal structure of human Hsp90 prompted the creation of its homology model for subsequent simulations. Two possible binding sites, C-terminal and N-terminal domains, were identified and considered in this study. Molecular docking followed by molecular dynamic simulations and post-dynamic analyses were performed to elaborate on the binding mechanism and relative binding affinities of nine FDA-approved HIV-1 protease inhibitors against human Hsp90. Our findings from thermodynamics calculations revealed that these inhibitors were more likely to bind to the N-terminal domain (~ 54.7 -83.03 kcal/mol) when compared to C-terminal domain. This appears to be the first account of a detailed computational investigation aimed to understand the binding mechanism of HIV

protease inhibitors binding to Hsp90. Information gained from this study should also provide a significant route map towards the design and optimisation of potential derivatives of protease inhibitors to treat breast cancer.

The results obtained will serve as a powerful tool in the drug design and development process. However, further experimental investigations will be useful to improve our computational findings.

DECLARATION 1 - PLAGIARISM

I, **<u>Sbongile H. Mbatha</u>**, declare that

The research reported in this thesis, except where otherwise indicated, is my original research.

This thesis has not been submitted for any degree or examination at any other university.

This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other resources have been quoted, then:

Their words have been re-written but the general information attributed to them has been referenced

Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

A detail contribution to publications that form part and/or include research presented in this thesis is stated.

Signed

………………………………………..

DECLARATION 2 – PUBLICATIONS

1. Mbatha, S.H. and Soliman, M.E. (2014) In-silico identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking Analysis and molecular dynamics simulations, *Combinatorial Chemistry & High Throughput Screening. (Accepted – proofs are enclosed)*

   Contribution:

   Mbatha, S. H.: contributed to the project by performing all the experimental work, data analysis and manuscript preparation.

   Soliman, M. E. S.: supervisor

2. Could the FDA-approved anti-HIV drugs be promising anticancer agents? Answer from extensive molecular dynamics analyses *(To be submitted for publication)*

   Contribution:

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

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

   Soliman, M.E.S.: supervisor

RESEARCH OUTPUT

A. PUBLICATIONS

1. Mbatha, S.H. and Soliman, M.E. (2014) In-silico identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking Analysis and molecular dynamics simulations, *Combinatorial Chemistry & High Throughput Screening.* (***Accepted – proofs are enclosed***)

2. Could the FDA-approved anti-HIV drugs be promising anticancer agents? Answer from extensive molecular dynamics analyses ***(Manuscript to be submitted for publication)***

B. CONFERENCES

1. Poster presentation "*In-silico* identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking Analysis and molecular dynamics simulations" - 3's company pharmaceutical conference, Cape Town, October 2013.

2. Oral presentation "*In-silico* identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking Analysis and molecular dynamics simulations" – HPC conference, Kruger National Park (Mpumalanga), December 2014.

ACKNOWLEDGEMENTS

I would like to thank God for His everlasting love an unfailing grace that kept me going even in hard times.

I would like to express my gratitude to the following people:

- My supervisor Prof. Mahmoud Soliman for his constant support and knowledge he provided me with in the course of study. I have received so much training and new skills in the medical research field.

- My colleagues in Molecular Modeling and Drug design field for their support.

- CHPC for their computational resources and support

- UKZN College of Health Sciences for financial support

- My family, more especially my late father for love and support I received and my friends for their support in all aspects of life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADP	Adenosine diphosphate
AKT	Serine/threonine kinase
Arg	Arginine
Asp	Aspartate
ATP	Adenosine diphosphate
CADD	Computer-aided drug design
CTD	C-terminal domain
Cys	Cysteine
DNA	Deoxyribonucleic acid
EPR	Enhanced permeability and retention
FDA	Food and Drug Administration
Gly	Glycine
GRAP94	Glucose-related protein 94
HER2	Human epidermal growth factor receptor 2
His	Histidine
HIV-1	Human Immunodeficiency Virus 1
HSF	Heat shock transcription factor
Hsp90	Heat shock protein 90
Ile	Isoleucine
Leu	Leucine
MD	Molecular dynamics
MDR	Multiple drug resistance
Met	Methionine
MHC	Major Histocompatibility complex
MM	Molecular mechanics
MM/GBSA	Molecular Mechanics General Born-Surface Area
MM/PBSA	Molecular Mechanics Poisson Boltzman-Surface Area
MMV	Molegro Molecular Viewer
NTD	N-terminal domain

PARP	Poly ADP (Adenosine Diphosphate)-Ribose Polymerase
PDB	Protein Data Bank
PDGF	Platelet-derived growth factor
PES	Potential energy surface
PI3K	Phosphoinositide 3-Kinase
PIs	Protease inhibitors
Pro	Proline
PSMA	Prostate-specific membrane antigens
RMSD	Root mean square deviation
RMSF	Root mean square fluctuations
ROS	Reactive oxygen species
SASA	Solvent accessible surface area
SBDD	Structure-based drug design
SE	Standard error
TPR	Tetratricopeptide repeat
TP53	Tumor protein 53
TRAP1	TNF (Tumor necrosis factor) receptor-associated protein 1
Trp	Tryptophan
Val	Valine
VEGF	Vascular endothelial growth factor

CHAPTER 1

1.1. Background and rationale for this study

A major problem in cancer research is the lack of new therapeutic strategies to control this disease. Cancer refers to an abnormal proliferation of cells, this being caused by multiple changes in gene expression, with the ability to invade neighboring tissues, thereby leading to death of the host (*1, 2*). Globally, cancer is regarded as a leading cause of death, constituting approximately 25% of all mortalities reported in Africa (*3*). The American Cancer Society reported that there is an expected increase in cancer incidences over the next 20 years due to an increase in the global population as well as aging (*4*). A better understanding of the pathological mechanism of cancer will assist in developing new anti-cancer drugs to reduce the disease burden. This thesis presents the investigation of anti-cancer drugs conducted on cancer-related enzyme targets, Cathepsin B and Hsp90, in order to gain insights into the discovery of potential anti-cancer drugs.

Cathepsins are lysosomal peptidases that are essential for intracellular proteolysis in mammalian cells. Cathepsin B is one of the lysosomal cysteine proteases that belong to the papain superfamily, and is involved in protein degradation (*5, 6*). Alterations in Cathepsin B expression, activity, localization and protein levels lead to conditions such as cardiovascular diseases, neurodegenerative disorder, inflammation and cancer (6). Recent studies have reported cathepsin B's involvement in cancer invasion and metastasis, making it a significant enzyme to target for cancer therapy (*7, 8*).

Heat Shock Protein (Hsp90) is involved in activating the signaling pathways of various client proteins that are involved in malignancies (*9*). Some of these client proteins include HER2 (promotes growth of cancer cells), AKT (activation of signaling pathways involved in cancer), EGFR (promotes solid tumor growth) and polymerase (directly involved in malignancies). Understanding the pharmacological inhibition of Hsp90 can provide a method of impeding various signaling pathways that are crucial for oncogenesis and malignant transformation (*10*). One approach used to reduce the drug discovery process and to advance cancer treatment is called "repositioning". This refers to a situation where a new agent has been shown to be effective in one disease, effort is devoted to further investigate its use in various combinations

for other disorders (*11*). In this study, the FDA-approved HIV-1 Protease Inhibitors (PIs) were investigated as potential anti-cancer drugs when interacting with Hsp90. These inhibitors have been shown to be involved in cancer inhibition, with the mechanisms being unclear.

1.2. Aims and objectives of this study

The major aim of this study was to investigate drugs as anticancer agents that target Cathepsin B and Hsp90.

Given the above general aim, this study has two practical aims:

1. **To discover novel Michael acceptor-type leads that form irreversible binding with Cathepsin B.**

To accomplish this, the following objectives were outlined:

1.1 To identify and acquire sets of Michael acceptor-type leads from the chemical database.

1.2 To identify new inhibitors that form covalent inhibition with Cathepsin B using covalent docking approach.

1.3 To describe the precise binding modes of covalent inhibitors.

1.4 To determine the compound stability on the Cathepsin B active site.

1.5 To determine active site amino acid contributions to ligand binding by performing the per-residue energy decomposition analysis.

2. To provide a molecular understanding of the effectiveness of the FDA-approved HIV-1 protease inhibitors (PIs) against Hsp90.

To accomplish this, the following objectives were outlined:

2.1 To create 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 homology model by plotting a Ramachandran plot to determine outliers.

2.3 To estimate the binding affinity of the protease inhibitors against the human Hsp90 homologue NTD and CTD.

2.4 To determine the inhibitor's stability on Hsp90 homologue

2.5 To estimate the contribution of each amino acids towards the overall binding by performing the per-residue decomposition analysis.

2.6 To perform a comparative study on which of the two Hsp90 terminals has better binding affinity for each drug.

1.3. Novelty and significance of this study

Despite ongoing research in anti-cancer drug development, cancer continues to receive a relatively low public health priority in Africa. Cancer-related morbidity and mortality rate are high and serious adverse effects of chemotherapy and radiation therapy contributes to an elevated disease burden, highlighting a need to discover and develop effective and safe anti-cancer drugs (*3, 4*). Cathepsin B and Hsp90 have been shown to be effective targets for cancer treatments, as they are involved in various pathologies and oncogenic processes in humans, making it important to gain insights into their interactions with inhibitors (*11, 12*). This research study serves as a cornerstone towards providing insights into the possible design and development of novel anti-cancer drugs.

1.4. Overview of this thesis

This thesis consists of the following six chapters (including the current chapter):

Chapter 1: In this chapter, background, aim, objectives and significance of the study are addressed and also provide with general outline and structure of the thesis.

Chapter 2: This chapter gives a historic background and an overview of cancer disease and its treatments, and then discusses the updated statistics on the diseases' global epidemiology. Important aspects of the proteins involved in the pathogenesis of cancer are also discussed in the concept of "the hallmarks of cancer" as described by Hannahan and Weinberg, which gives a clear description of the remarkable diversity of neoplastic diseases. A brief discussion on targeted therapy is presented and the essential drug targets, cathepsin B and Hsp90, selected as

the main focus of this research, are described in detail. This includes: their detailed structures (including their active sites) and functions; their pathological pathways that contribute to cancer, and also their general mechanism of action.

Chapter 3: This chapter presents an introduction to, and application of computational chemistry, different molecular modeling and molecular simulation techniques. An overview of the theoretical descriptions of the computational methods is provided and a description of several computational tools used in this study then follows. These methods include: molecular dynamics (MD) simulations, molecular mechanics (MM), molecular docking and binding free energy calculations.

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

This research paper is titled "In-silico identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking analysis and molecular dynamics simulations.", and was published in the Journal of Combinatorial Chemistry & High Throughput Screening. It addresses Aim 1 and Objectives 1.1 to 1.5 of this study.

The chapter presents a comparative study of Cathepsin B and its inhibitors with several computational techniques being adopted for optimising the results. A 'ZINC' database search for compounds with Michael-acceptor like features was carried out. Covalent docking was conducted for compounds that showed better binding to Cathepsin B. The final 'hit' compound with the highest covalent binding affinity was selected for further analysis using MD simulations. To verify the docking results, binding free energy calculations were carried out. Per-residue decomposition analysis was also performed to determine the most interactive amino acid residues within the active site.

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 titled "Could the FDA-approved anti-HIV drugs be promising anticancer agents?" and it addresses Aim 2 and Objectives 2.1 to 2.6.

It presents the comparative study on Hsp90 with FDA-approved HIV-1 protease inhibitors that have shown inhibitory effects on cancer. Several computational methods were adopted for optimising the results. Due to the absence of the complete human Hsp90 crystal structure, a model from the protein sequence was built using the relevant crystal structure with PDB Codes: 2CG9, 3PRY and 3HJC as reference templates. Molecular docking was carried out on all the FDA-approved PIs to build Hsp90/HIV-1 PIs complexes. MD simulations to determine the compounds stability within a human Hsp90 homologue active site and binding free energy calculations were also performed. Per-residue decomposition analysis was also performed to determine the most interactive amino acid residues within the active site. Other post-dynamic calculations were performed on both domains of human Hsp90 homologue.

Chapter 6: This chapter answers the research question, and addresses the study aim by reviewing the aim findings from the two sub-aims. It presents the study limitations, recommendations and suggestions for future research.

References

1. de Mesquita, M. L., de Paula, J. E., Pessoa, C., de Moraes, M. O., Costa-Lotufo, L. V., Grougnet, R., Michel, S., Tillequin, F., and Espindola, L. S. (2009) Cytotoxic activity of Brazilian Cerrado plants used in traditional medicine against cancer cell lines, *J. Ethnopharmacol. 123*, 439-445.

2. Vineis, P., and Wild, C. P. (2014) Global cancer patterns: causes and prevention, *Lancet 383*, 549-557.

3. 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.

4. Unnati, S., Ripal, S., Sanjeev, A., and Niyati, A. (2013) Novel anticancer agents from plant sources, *Chinese Journal of Natural Medicines 11*, 16-23.

5. Sosic, I., Mirkovic, B., Turk, S., Stefane, B., Kos, J., and Gobec, S. (2011) Discovery and kinetic evaluation of 6-substituted 4-benzylthio-1,3, 5-triazin-2(1H)-ones as inhibitors of cathepsin B, *Eur. J. Med. Chem. 46*, 4648-4656.

6. Tomoo, K. (2010) Development of Cathepsin Inhibitors and Structure-Based Design of Cathepsin B-Specific Inhibitor, *Current Topics in Medicinal Chemistry 10*, 696-707.

7. Beckham, T. H., Lu, P., Cheng, J. C., Zhao, D., Turner, L. S., Zhang, X. Y., Hoffman, S., Armeson, K. E., Liu, A. G., Marrison, T., Hannun, Y. A., and Liu, X. (2012) Acid ceramidase-mediated production of sphingosine 1-phosphate promotes prostate cancer invasion through upregulation of cathepsin B, *International Journal of Cancer 131*, 2034-2043.

8. Gondi, C. S., and Rao, J. S. (2013) Cathepsin B as a cancer target, *Expert Opin. Ther. Targets 17*, 281-291.

9. Mahalingam, D., Swords, R., Carew, J. S., Nawrocki, S. T., Bhalla, K., and Giles, F. J. (2009) Targeting HSP90 for cancer therapy, *Br. J. Cancer 100*, 1523-1529.

10. Sankhala, K. K., Mita, M. M., Mita, A. C., and Takimoto, C. H. (2011) Heat Shock Proteins: A Potential Anticancer Target, *Curr. Drug Targets 12*, 2001-2008.

11. Nomura, T., and Katunuma, N. (2005) Involvement of cathepsins in the invasion, metastasis and proliferation of cancer cells, *The journal of medical investigation : JMI 52*, 1-9.

12.   Sidera, K., Gaitanou, M., Stellas, D., Matsas, R., and Patsavoudi, E. (2008) A critical role for HSP90 in cancer cell invasion involves interaction with the extracellular domain of HER-2, *J. Biol. Chem. 283*, 2031-2041.

CHAPTER 2

2. Introduction to Cancer and its drug targets

This chapter briefly provides an overview on the historical background of the diagnosis and treatments of cancer. Problems encountered during chemotherapeutic drug development are briefly discussed. It then provides insights into the structural and functional properties of the different cancerous drug targets used in this research.

2.1. Background

Cancer is ranked as the second highest cause of death worldwide, the increasing burden of cancer in developing countries is attributed to numerous factors, such as population aging and growth and, cancer-associated lifestyle factors (smoking, physical inactivity, and unhealthy diets) (*1-3*). In people redisposed to the negative consequences of these factors, cancers can manifest in various forms, including tumours, leukemia, sarcomas, Hodgkin disease and non-Hodgkin lymphoma.

Malignant tumors are cancerous growths that invade nearby tissues and can spread to other parts of the body, with example of tumors including the brain, lungs, breast, prostate, skin, and colon (*4*). Breast cancer is the most commonly diagnosed type of cancer globally, with an expected 20% increase in related deaths in the next 10 years (*5-7*). Global increases in newly diagnosed cancer cases and deaths remain unchecked (*5*).

There has been a paradigm shift in drug development, from empirical cytotoxic approach identified by screening a variety of natural and synthetic compounds, to therapies that act at specific molecular targets (*8-10*). In recent decades, an increased understanding of genetics and molecular biology of cancer has facilitated the identification of potential molecular targets for anticancer drug discovery and development (*11, 12*). The main purpose of the target-based drug discovery approach is to develop agents that block the pathogenic mechanism responsible for malignant transformation (*13*), as well as to discover etiologic agents that demonstrate minimal adverse effects on normal cells (*9*). However, for the majority of commonly known classified cancers, there are no specific etiologic agents available, thereby prompting further investigation (*9*). Hanahan and Weinberg described the functional, molecular and biological traits that

differentiate tumor from normal cells as the 'hallmarks of cancer' (*14, 15*). This understanding has resulted in the development of several compounds to treat advanced malignancies with specific molecular targets (*14, 16*).

2.2. The Hallmarks of Cancer

Ongoing cancer research has not only enriched understanding the complexity of this disease, but has also revealed the dynamic alterations in a cancerous genome (*15*). Despite the rapid advancements in research, it remains difficult to develop effective drugs due to the complexity of cancer cells and the tumor microenvironment (*14, 17*). Hannahan and Weinberg's (*15, 18*) concept of the 'hallmarks of cancer', provides with deeper insights into the traits of remarkable diversity of neoplastic diseases and have assisted with developing cancer drugs for particular molecular targets that incorporate these hallmarks (*19*). Figure 2.1 provides some examples of drugs that have been developed to target 'the hallmarks of cancer'. In addition, the 'hallmarks of cancer' have served as a powerful guide for translational research that aims to improve and develop early detection, screening, and treatment, and to enhance the quality of life and well-being of cancer patients (*20*).

Hannahan and Weinberg indicate that: a. normal cells progressing towards a neoplastic state acquire progression of several hallmark abilities, and b. the multistep process of human tumor pathogenesis may be rationalized by emerging cancer cells to acquire cancerous traits that facilitate tumorigenesis and eventually malignancy (*15, 18*). The hallmarks or traits are mainly the acquired capabilities that enable cancer cells survival, proliferation and spread (*17, 18*). These hallmarks include sustaining proliferative signaling, insensitivity to growth suppressors, an ability to evade apoptosis, limitless replicative potential, sustained angiogenesis and an ability to metastasize (*15, 18, 19*). Progresses in the past decade have included four biological capabilities, namely, genome instability, tumour-promoting inflammation, reprogramming of cell bioenergetics and an ability to escape immune destruction (*18, 20*). These hallmarks are discussed in detail below.
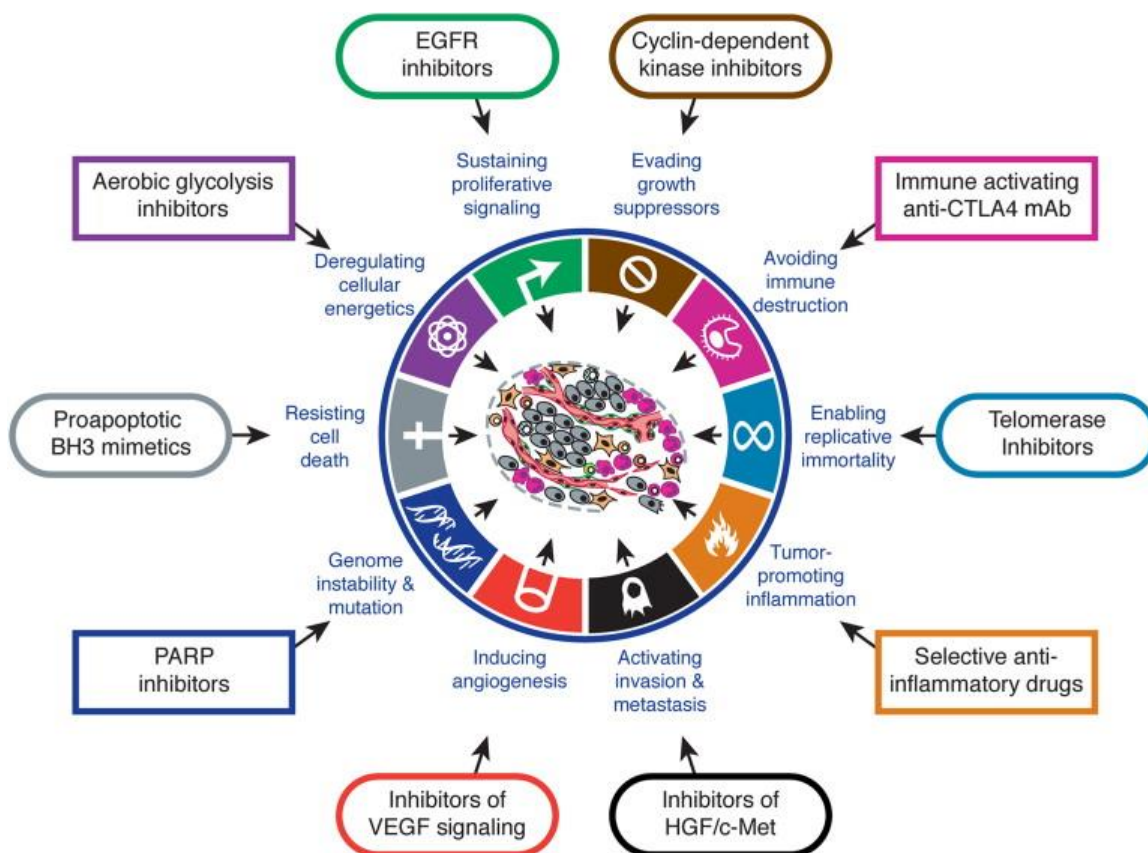
Figure 2.1: The hallmarks of cancer with examples of therapeutics targeting each hallmark (*18*).

2.2.1. Sustaining proliferative signaling

During growth and differentiation of normal cells, cell proliferation is controlled to ensure that a homeostatic cell number is achieved, thereby resulting to maintaining a normal tissue architecture and function (*18*). Normal cells require mitogenic growth signals to move from a state of rest to a state of active proliferation, which is generally produced by neighboring cells (*14, 15*). Transmembrane receptors transmit signals into the cells by binding to different classes of signaling molecules such as cell-to-cell adhesion interaction molecules, diffusible factors and extracellular matrix component (*15*). Studies have shown that without such stimulatory signals, proliferation of any type of normal cells cannot occur (*14, 15, 21*). However, tumor cells continue to proliferate even in the absence of stimulatory signals by mimicking normal growth signaling (39). These imitated signals are produced by oncogens that are involved in most of the cellular signaling pathways (*14, 18, 22, 23*). One of the signals produced by oncogens is the

phosphatidylinositol-3-kinases (PI3K)/AKT pathway, which is responsible for promoting cell growth, proliferation, survival and migration (*14, 24, 25*). Inhibiting the biological function of (PI3K)/AKT and the self-sufficiency of its growth signals has shown important effects in cancer regimens (*26*).

2.2.2. Insensitivity to growth suppressors

Apart from the hallmark ability to sustain proliferative signals, cancer cells have also mastered the ability to evade powerful mechanisms that negatively regulate their proliferation (3). In normal tissues, several growth arrest signals function to prevent their proliferation, thus maintaining cellular quiescence and homeostasis (*15, 18*). Limitations in cell growth and proliferation are achieved by tumor suppressor genes that encode the retinoblastoma (RB) family of proteins and cyclin-dependant kinases (CDKs) (*14, 27, 28*). Cancer cells evade RB signals, either by gene mutation or altered expression of cyclin-dependant kinases (*14, 26*). As a result, CDK inhibitors that regulate cyclin function are often found in malignancy (*26*).

2.2.3. Resisting apoptosis

Normal cells do not only maintain cell homeostasis through proliferation but also through apoptosis, which is a process of programmed cell death that takes place in order to control the organism's cell growth and development (*14, 29, 30*). Alternatively, tumor cells have evolved numerous strategies to resist or circumvent apoptosis (*14, 31*). The most common strategy in which tumor cells evade apoptosis is through the loss of the TP53 tumor suppressor gene function. This encodes the p53 tumor suppressor gene responsible for regulating the cell cycle, which then functions as a tumor suppressor (*18*). Tumor cells also increase the expression of antiapoptotic regulators, namely, Bcl2 and Bcl-xL, by the downregulation of pro-apoptotic factors such as Bax, Bim and Puma (*18, 31*). Various approaches to reversing apoptosis are currently under development and are aimed at magnifying chemotherapeutic effects (*26, 32*).

2.2.4. Limitless replicative potential

Numerous studies have demonstrated that normal cells carry an intrinsic, stand-alone program that allows them to limit their proliferation and multiplication (*14, 15*). Some cell populations enter a "senescence" state (alive but non-proliferative state) after undergoing a certain number of cell divisions or a "crisis" leading to cell death (*14, 18, 33*). Rare cells that survive crisis show unlimited replicative potential, thus becoming immortilised (a characteristic of established cell

line due to their ability to proliferate without evidence of senescence and crisis) (*15, 20*). Telomerase, discovered by Greider and Blockburn, is an enzyme that synthesizes DNA telomeres and maintains DNA molecule stability (*14, 33*). In the absence of this enzyme telomerase becomes shorter, thus preventing DNA replication leading to cell cycle arrest or death (*14, 34*). Tumor cell, on the other hand, enable replicative immortality by two mechanisms: a. by an increase in the expression of telomerase, and b. by activating an alternative recombination-based telomere maintenance mechanism, which is less common (*14, 18, 19*). As a result, drugs targeting telomerase are said to have an increased specificity and a reduced toxicity to normal tissue (*26, 35*).

2.2.5. Inducing angiogenesis

During embryogenesis, new blood vessels are developed from the existing ones (angiogenesis) to supply oxygen and nutrients that are crucial for cell function and survival (*15, 18*). Once the new tissues are formed, the process of angiogenesis is adequately and transitory regulated (*14, 15, 18*). In contrast, tumor cells require angiogenesis early during the multi-step development of invasive cancers, thereby ensuring that its oxygen and nutrients needs are met (*15, 18, 26*). Tumor vasculature is characterized by enlarged vessels, microhemorrhaging, irregular blood flow, leakiness, excessive vessel branching and an abnormal level of endothelial cell proliferation and apoptosis (*18, 36, 37*). Angiogenesis-induced progression of micrometastases to lethal macrometastases is said to be the major cause of death in cancer patients (*19*). Drugs targeting angiogenesis have drastically improved in the last 30 years and most of these drugs inhibits pro-angiogenic factors such as VEGF, EGF, PDGF and interleukins (*14, 19, 38, 39*).

2.2.6. Activating invasion and metastasis

Metastasis is a process by which tumor cells move away from the primary organ, invade adjacent tissues and travel to distant sites, via hematogenous or lymphatic circulation, where they succeed in finding new colonies (*15, 40*). This acquired capability enables tumor cells to invade and metastasize to new terrain where there are sufficient space and nutrients (*15*). This is due to alterations in the cancer cell shape and their clinging to other cells or to the extracellular matrix (ECM) (*18*). E-cadherins, a key cell-to-cell adhesion molecule, is downregulated in human carcinomas, thus cell quiescence cannot be maintained (*18, 26*). Most drugs created to inhibit

metastasis are currently in clinical trials, with none having been approved by USA Food and Drug Administration (FDA) (*14*).

2.2.7. Genome instability and mutation

The above-mentioned hallmarks depend mainly on the successions of genomic alterations of neoplastic cells (*18*). For example, the acquired capability of evading growth suppressors is solely dependent on DNA methylation and histone modifications (*41, 42*). However, some genomic alterations are triggered by nonmutational alterations that affect the regulation of gene expression (*43*). "Housekeeping" or "maintenance" genes are those linked to genetic instability development, and are responsible for detecting, preventing and repairing DNA damage (*14, 44*). One of the well-studied genes that senses single stranded DNA damage and aids in its repair in neoplastic cells is the PARP1 gene (*45*). Drugs known as chemotherapy-enhancing agents have been developed for PARP inhibition (*26, 45-47*).

2.2.8. Tumor-promoting inflammation

Progress in hallmark cancer research has demonstrated that tumor-associated inflammatory response has the unexpected effect of promoting tumorogenesis and progression on developing neoplasia (*18*). It has been stated that inflammation play essential role to numerous hallmark capabilities by providing bioactive molecules to the tumor microenvironment (*48*). These include, invasion and metastasis, survival factors that limit cell death, extracellular matrix-modifying enzymes for angiogenesis, proangiogenic factors, and growth factors that enables proliferative signaling (*26, 49*). In addition, inflammatory cells also release actively mutagenic chemicals (reactive oxygen species, ROS) to nearby tissues, resulting in intensified malignancy (*18, 49*).

2.2.9. Reprogramming energy metabolism

Under aerobic conditions, normal cells undergo the process called glycolysis where glucose is converted first to pyruvate in the cytosol and then to carbon dioxide in the mitochondria under anaerobic conditions (*18*). This is different to cancer cells, which limit their metabolism to glycolysis in the process called the "Warburg effect" (*50, 51*). The discovery by Warburg was that: even in the presence of oxygen (aerobic condition), cancer cells constantly enhance their glucose metabolism, and that this "aerobic glycolysis" involves alterations in ATP production by oxidative phosphorylation even in quiescent and differentiated cells (*14, 51, 52*). Tumor cells

depend on glycolysis for cell growth and division to ensure that all their metabolic requirements are met (*18, 53*). This hyperglycolytic state of tumor cells has shifted the focus of anticancer drug development into drugs that inhibit glycolysis mediated by 2-deoxy-D-glucose (2DG) (*53*).

2.2.10. Evading immune destruction

The theory of immune surveillance suggests that the immune system recognizes and eliminates numerous incipient cancer cells (*18, 54*). Solid tumors that develop have the ability to limit this immune recognition, thereby evading eradication (*55*). Strategies that enable tumor cells to evade immune destruction are: promoting cell recognition errors by Natural Killer cells (NK), physical exclusion from the immune cells at the tumor site, and adapting poor immunogenicity by reducing major histocompatibility complex (MHC) expression (*14, 56*). Passive immunization, monoclonal antibodies and cytokine administration have been introduced to activate antitumor immunity from the observation that a wide variety of proteins expressed by tumors can be recognized by the immune system (*57*).

2.3. Targeted therapy

Although the principal options for cancer regimens (i.e. surgery, radiotherapy and chemotherapy) have proved successful, the beneficial effects pose several health risks to these patients (*58, 59*). Some of these include: ineffective therapeutic drug concentrations to tumor tissues, high dose requirements, life threatening side-effects due to non-specificity of anticancer agents and multiple drug resistance (MDR) development that triggers cross-resistance (*59-61*). A deeper understanding of molecular biology and genetics of cancer has allowed for revolutionary changes in drug development tactics, whereby novel cancer drugs have been developed to interact with specific molecular targets (*61-64*). Targeted therapy enhances patient survival rate due to the high therapeutic index and a dose below the maximal threshold required for optimally effective treatments (*60-62*). Peptides, monoclonal antibodies, vitamins (folic acid, aptamers, lectins), hormones and growth factors are examples of compounds utilized as targeting moieties in targeted cancer therapy (*58, 65*). The basic principle of drug targeting involves classifying these substances by their method of reaching the targeted tissue, leading to the distinction between passively and actively targeted therapy (*42, 58*) as discussed below:

a. Passive Targeting

Passive targeting is a mechanism where drug or drug carrier system accumulate at a specific site due to physicochemical or pharmacological factors in malignant tissues (*42*). It relies mainly on the anatomical and functional differences in vasculature and permeability between normal and tumor tissues. As stated previously, malignant tissues show a generally heterogeneous distribution, have a high vascular density and larger in size, and are more permeable and leaky compared to the normal tissues (*42, 58*). This leaky vasculature enables extravasation of nanocarriers to the tumors. In addition, the impaired lymphatic channels in tumor tissue enhance the accumulation and retention of high molecular drugs, as well as the availability of chemotherapy for cell death (*42, 58*), this phenomenon being referred to as the enhanced permeability and retention (EPR) effect.

b. Active Targeting

Active targeting involves the expression of receptor and antigen by cancerous cells as compared to normal cells (*42*). For example, folate receptors and some surface bound antigens, such as prostate-specific membrane antigens (PSMA), which are uniquely and highly expressed in cancerous cells compared to normal cells. Active targeting employs a targeting moiety called a ligand to achieve surface conjugation of nanoparticles with particular targets, thereby increasing the specificity and affinity for surface receptor or antigen expression on specific tissues (*42, 58*). It has been stated that, active targeting improves drug delivery localization to the tumor and aids in drug delivery system uptake by endocytosis (*42*). Some examples include, aptamers, peptides, sugar molecules, antibodies and nanobodies (*42, 58*).

2.4. Cancer drug targets

While there are a number of cancer drug targets, only those that were used in this study, due to their involvement in various pathologies and oncogenic processes in humans, are discussed. These include the two drug targets, namely: Cathepsin B and Heat shock protein 90 (Hsp90), each of which will be reviewed with respect to their structure, active site and mechanism of action.

2.4.1. Cathepsin B

Cathepsin B is the most abundant lysosomal cysteine protease that belongs to the papain family of enzymes and it is involved mainly in intracellular protein degradation (*66*). It is also involved

in a variety of physiological and pathological processes such as bone resorption, antigen processing, cartilage proteoglycan breakdown, intracellular protein turnover and malignant progression of tumors (*67-69*). Cathepsin B is synthesized as an inactive proenzyme, which is then translocated to the lysosomes for modification and maturation into active enzyme (*70*). A mature Cathepsin B exhibits optimal activity at slightly acidic pH and it becomes inactivated at neutral or slightly alkaline pH due to irreversible denaturation (*66, 69*). A remarkable feature of Cathepsin B that makes it distinct from other cysteine peptidases is its ability to acts both as an endo- and an exopeptidase (peptidyldipeptidase) (*66, 67*).

Recent studies have shown that Cathepsin B is associated with invasive and metastatic phenotypes in cancers. This association is reported to be a result of the abnormal regulation of Cathepsin B causing cells to acquire an oncogenic characteristic (*71, 72*). In addition, overexpression of Cathepsin B has been observed in various malignancies including breast, lung, prostate, colorectal and brain cancer (*71*). As a result, targeting Cathepsin B will have significant clinical relevance in cancer therapy (*73*).

2.4.1.1. Structure of Cathepsin B

The matured form of Cathepsin B is a 30kDa bilobal protein comprised of both one light and heavy chains. The light chain comprises residues Leu1-Asp47 from the amino acid terminal and the heavy chain comprises Val50-Asp254 from the carboxyl terminal (*69*). Located between the two lobes are the active site and the substrate binding cleft (**Figure 2.2**). In addition, human Cathepsin B also possesses six disulfide bridges that are confined to half the amino-terminal of the molecule (*66*).

Numerous X-ray crystallographic studies have shown that Cathepsin B is unique amongst other cysteine peptides, as it consists of a flexible structural element known as the "occluding loop", which consists of Ile105-Pro126 amino acid residues. The open and closed conformations of the loop enable the endo- and exopeptidase characteristics of Cathepsin B, respectively (*67, 68*).
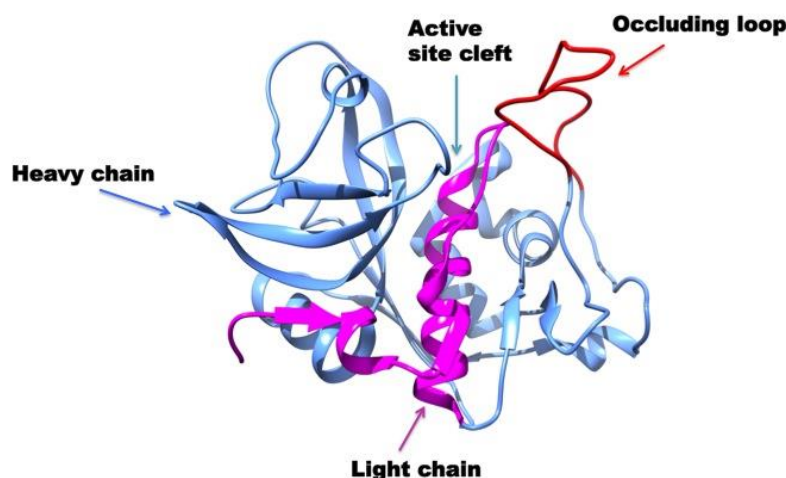
Figure 2.2: The structure of Cathepsin B showing the occluding loop (red), the active site cleft within the light chain (pink) and heavy chain (blue).

2.4.1.2.   Cathepsin B active site

As illustrated in **Figure 2.3**, the crucial active site residues of Cathepsin B are Cysteine 29 (Cys29) and Histidine 199 (His199). Cysteine 29 is located in the middle region within the occluding loop that catalyses the peptide bond cleavage on the left lobe, which then interacts with His199 on the right lobe (*69*). It has been reported that the Cys29 thiole and imidazole side chains and His199 form an ion pair in a pH ranging from 4.0 to 8.5, respectively (*74*). Furthermore, the occluding loop residues, His110 and His111, regulate the substrate access to Cathepsin B active site which then binds to the carboxylic group of the C terminal of the substrate. These residues, His110 and His111, are also responsible for Cathepsin B stabilization (*67*).
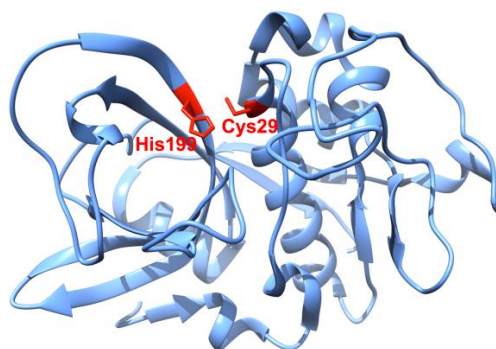

Figure 2.3: The active site cleft of Cathepsin B showing Cys29 and His199 amino acid residues in red.

17

The dual mechanism of Cathepsin B is its ability to exhibit both endopeptidase and exopeptidase activities. When acting as an endopeptidase, Cathepsin B cleaves the internal peptide bond through the interaction between Cys29 and His199 with a broad specificity and preference of basic residues or phenylalanine (*66, 74*). Cathepsin B accepts an arginine due to the availability of a glutamic acid (E245) that is situated at the top of the binding pocket (*74*). When acting as an exopeptidase (peptidyldipeptidase), Cathepsin B removes dipeptides from the C-terminus of proteins and peptides. This activity occurs at a low pH which holds the occluding loop at a closed conformity at the substrate binding cleft (*68*). Consequently, the end of the active site cleft is partially blocked by the loop, thereby allowing the positively charged imidazole group (His111) to accept a negative charge associated with the C-terminal (*74*).

2.4.2. Heat shock protein 90 (Hsp90)

Heat shock proteins (HSPs) are highly conserved molecular chaperones that play essential roles in protein folding, intracellular disposition and proteolytic turnover in response to stress, thereby serving as central integrators of cellular homeostasis (*75-77*). These molecules were first discovered in 1962 in *Drosophila melanogaster* as proteins stimulated by thermal stress, and then later shown to be excited by other stressful cellular conditions, such as hypoxia, inflammation, radiation and toxins. HSP regulation occurs through the action of heat shock transcription factors (HSF), with mammalian forms being classified into five families according to their molecular weight: Hsp40, Hsp60, Hsp70, Hsp90 and a small HSP such as Hsp27 (*76*). Among these families of chaperoning proteins, Hsp90 has been found to be the most abundant cellular chaperon protein.

Heat shock protein 90 (Hsp90) is an ATP-dependent molecular chaperone that is crucial for maintaining, activating and stabilizing numerous client proteins in eukaryotes. Most of these client proteins are involved in signaling pathways such as signal transduction, protein degradation and intracellular transport (*78, 79*). Hsp90 client proteins include HER2, AKT, EGFR, SRC, CRAF, ABL, FLT3, MET, BRAF, telomerase (directly involved in malignancy), androgen and estrogen receptor and hypoxia-inducible factor ((HIF)-1α). However, there is a variety of client proteins that have been identified and the list is continuously updated (*76*). To achieve its function, Hsp90 form a multicomponent complex with a large group of co-factors, known as co-chaperones, to facilitate recognition and activation of client proteins (*78*). This

Hsp90/co-chaperone complex aids regulation of ATPase conformational changes. With the identified co-chaperones including Cdc37, Hsp40, Hsp70, p23 and Hop (Hsp40 and Hsp70 are organizing proteins) (*80, 81*). As a result, Hsp90 has appeared as an interesting target for cancer therapy due to its involvement in activating various signaling pathways involved in malignancy.

2.4.2.1.   Structure of Hsp90

Hsp90 exists in all eukaryotic cells and consists of four members: the stress-induced Hsp90α, Hsp90β that is constitutively expressed, glucose-related protein 94 (GRP94) located in the endoplasmic reticulum and mitochondrial TNF receptor-associated protein 1 (TRAP1) (*81*). Structurally, Hsp90 is a homodimer consisting of three flexibly linked domains: firstly, an N-terminal ATP-binding domain (NTD or ND), where ATP binding and hydrolysis take place for the chaperoning activity of Hsp90 (refer to **Figure 2.4**). The ATP binding at this site constitutes to the conformational dynamic state of Hsp90, and also controls its interaction with client proteins and chaperones (*75, 82, 83*). Secondly, is the middle domain (MD), which contributes to client protein and chaperone binding to Hsp90 (*77, 78, 83*). Moreover, there is a highly charged linker that connects the NTD to the MD, which is 80 residues long in human Hsp90β (*75*). Lastly, the C-terminal dimerization domain (CTD or CD) is involved in Hsp90 dimerization and is also reported to contain an MEEVD motif required for TPR binding (*78, 80, 83*).
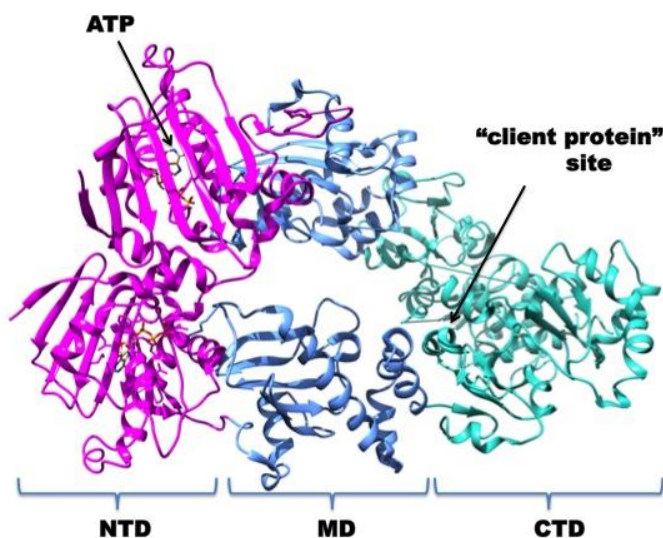

Figure 2.4: The structure of Hsp90 (Hsp90α and Hsp90β both with ATP), showing the NTD (pink), MD (blue) and the CTD (cyan).

2.4.2.2.   Hsp90 active site

The NTD entails an ATP binding site composed of the well-conserved residues Ser108, Phe133, Asp97, Cys53 and Met93 for Hsp90 ATPase activity (*84*). Most of the pharmaceutical studies have focused on developing drugs that inhibit ATPase activity for cancer therapy, an approach that has shown some success (*85, 86*). The MD is reported to contain catalytic residue Arg391 that is essential for forming ATPase site, and is also involved in ATP hydrolysis (*78, 81, 83*). The CTD is composed of the well conserved residues namely, Cys590, Gly540, Trp598, Thr537, Ser594 and Ile590, thereby permitting the CTD dimerization activity. In addition, while the far end of the CTD was implicated as an alternative ATP binding site, there is no structural evidence for this suggestion (*78, 83*).

2.4.2.3.   Conformational dynamics of Hsp90

ATP binding to Hsp90 triggers conformational changes that are required to facilitate the chaperoning activity of Hsp90. This forms a multicomplex structure with Hsp40, Hsp70, Hop and Cdc37 co-chaperones, thereby aiding client protein recognition (*81*). In this 'open conformation' state, Hsp90 becomes a mature complex, ready to perform its functions of client protein folding and stabilization (*81*). As mentioned earlier, the MD interacts with the co-chaperone aha1, and contributes to ATP hydrolysis to ADP and p23 binding. Consequently, the binding of the p23 co-chaperone to the Hsp90/client protein complex leads to client protein refolding and release (*76, 81*). Hsp90 then reaches a 'closed conformation' state, where N-terminal domains are dimerized (**Figure 2.5**). The released Hsp90 and co-chaperones will participate in the next ATPase cycle, while client proteins are degraded by ubiquitin proteasomes (*76*).
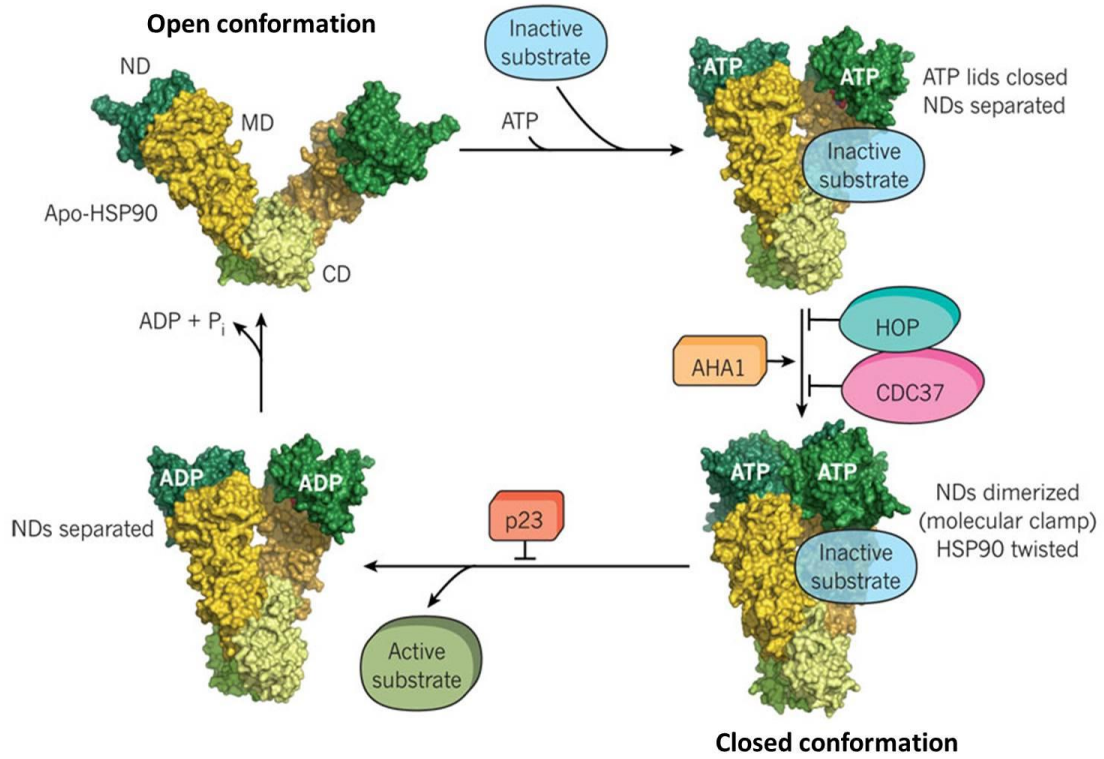
Figure 2.5: General mechanism of the ATPase cycle of Hsp90 showing an open and a closed conformation adapted during the cycle (*87*).

References

1.   Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., Sung, B. Y., and Aggarwal, B. B. (2008) Cancer is a Preventable Disease that Requires Major Lifestyle Changes, *Pharm. Res. 25*, 2097-2116.

2.   Vedham, V., Divi, R. L., Starks, V. L., and Verma, M. (2014) Multiple Infections and Cancer: Implications in Epidemiology, *Technol. Cancer Res. Treat. 13*, 177-194.

3.   Vineis, P., and Wild, C. P. (2014) Global cancer patterns: causes and prevention, *Lancet 383*, 549-557.

4.   Carroll, K. K. (1998) Obesity as a risk factor for certain types of cancer, *Lipids 33*, 1055-1059.

5.   Are, C., Rajaram, S., Are, M., Raj, H., Anderson, B. O., Swamy, R. C., Vijayakumar, M., Song, T., Pandey, M., Edney, J. A., and Cazap, E. L. (2013) A review of global cancer burden: Trends, challenges, strategies, and a role for surgeons, *Journal of Surgical Oncology 107*, 221-226.

6.   DeVita, V. T., and Rosenberg, S. A. (2012) Two Hundred Years of Cancer Research, *N. Engl. J. Med. 366*, 2207-2214.

7.   Patel, J. D., Krilov, L., Adams, S., Aghajanian, C., Basch, E., Brose, M. S., Carroll, W. L., de Lima, M., Gilbert, M. R., Kris, M. G., Marshall, J. L., Masters, G. A., O'Day, S. J., Polite, B., Schwartz, G. K., Sharma, S., Thompson, I., Vogelzang, N. J., and Roth, B. J. (2014) Clinical Cancer Advances 2013: Annual Report on Progress Against Cancer From the American Society of Clinical Oncology, *Journal of Clinical Oncology 32*, 129-+.

8.   Gelmon, K. A., Eisenhauer, E. A., Harris, A. L., Ratain, M. J., and Workman, P. (1999) Anticancer agents targeting signaling molecules and cancer cell environment: Challenges for drug development?, *Journal of the National Cancer Institute 91*, 1281-1287.

9.   Giang, I., Boland, E. L., and Poon, G. M. K. (2014) Prodrug applications for targeted cancer therapy, *The AAPS journal 16*, 899-913.

10.   Kummar, S., Gutierrez, M., Doroshow, J. H., and Murgo, A. J. (2006) Drug development in oncology: classical cytotoxics and molecularly targeted agents, *Br. J. Clin. Pharmacol. 62*, 15-26.

11.   Geromichalos, G. D. (2007) Importance of molecular computer modeling in anticancer drug development, *J. Buon 12*, S101-S118.

12.   Saijo, N., Tamura, T., and Nishio, K. (2003) Strategy for the development of novel anticancer drugs, *Cancer Chemother. Pharmacol. 52*, S97-S101.

13.   Fox, E., Curt, G. A., and Balis, F. M. (2002) Clinical trial design for target-based therapy, *Oncologist 7*, 401-409.

14.   Bailon-Moscoso, N., Romero-Benavides, J. C., and Ostrosky-Wegman, P. (2014) Development of anticancer drugs based on the hallmarks of tumor cells, *Tumor Biol. 35*, 3981-3995.

15.   Hanahan, D., and Weinberg, R. A. (2000) The hallmarks of cancer, *Cell 100*, 57-70.

16.   Saijo, N., Tamura, T., and Nishio, K. (2000) Problems in the development of target-based drugs, *Cancer Chemother. Pharmacol. 46*, S43-S45.

17.   Floor, S. L., Dumont, J. E., Maenhaut, C., and Raspe, E. (2012) Hallmarks of cancer: of all cancer cells, all the time?, *Trends Mol. Med 18*, 509-515.

18.   Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of Cancer: The Next Generation, *Cell 144*, 646-674.

19.   Siddiqa, A., and Marciniak, R. (2008) Targeting the hallmarks of cancer, *Cancer Biol. Ther. 7*, 740-741.

20.   Hainaut, P., and Plymoth, A. (2013) Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy, *Curr. Opin. Oncol. 25*, 50-51.

21.   Engelman, J. A. (2009) Targeting PI3K signalling in cancer: opportunities, challenges and limitations, *Nat. Rev. Cancer 9*, 550-562.

22.   Bhowmick, N. A., Neilson, E. G., and Moses, H. L. (2004) Stromal fibroblasts in cancer initiation and progression, *Nature 432*, 332-337.

23.   Cheng, N., Chytil, A., Shyr, Y., Joly, A., and Moses, H. L. (2008) Transforming Growth Factor-beta Signaling-Deficient Fibroblasts Enhance Hepatocyte Growth Factor Signaling in Mammary Carcinoma Cells to Promote Scattering and Invasion, *Mol. Cancer Res. 6*, 1521-1533.

24.   Maira, S. M., Stauffer, F., Brueggen, J., Furet, P., Schnell, C., Fritsch, C., Brachmann, S., Chene, P., De Pover, A., Schoemaker, K., Fabbro, D., Gabriel, D., Simonen, M., Murphy, L., Finan, P., Sellers, W., and Garcia-Echeverria, C. (2008) Identification and characterization of NVP-BEZ235, a new orally available dual phosphatidylinositol 3-

kinase/mammalian target of rapamycin inhibitor with potent in vivo antitumor activity, *Mol. Cancer Ther. 7*, 1851-1863.

25.   Zaytseva, Y. Y., Valentino, J. D., Gulhati, P., and Evers, B. M. (2012) mTOR inhibitors in cancer therapy, *Cancer Lett. 319*, 1-7.

26.   Sledge, G. W., and Miller, K. D. (2003) Exploiting the hallmarks of cancer: the future conquest of breast cancer, *Eur. J. Cancer 39*, 1668-1675.

27.   Sherr, C. J. (1996) Cancer cell cycles, *Science 274*, 1672-1677.

28.   Vermeulen, K., Van Bockstaele, D. R., and Berneman, Z. N. (2003) The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer, *Cell Prolif. 36*, 131-149.

29.   Kroemer, G., Galluzzi, L., Vandenabeele, P., Abrams, J., Alnemri, E. S., Baehrecke, E. H., Blagosklonny, M. V., El-Deiry, W. S., Golstein, P., Green, D. R., Hengartner, M., Knight, R. A., Kumar, S., Lipton, S. A., Malorni, W., Nunez, G., Peter, M. E., Tschopp, J., Yuan, J., Piacentini, M., Zhivotovsky, B., and Melino, G. (2009) Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009, *Cell Death Differ. 16*, 3-11.

30.   Thompson, C. B. (1995) Apoptosis in the pathogenesis and treatment of disease, *Science 267*, 1456-1462.

31.   Elmore, S. (2007) Apoptosis: A review of programmed cell death, *Toxicol. Pathol. 35*, 495-516.

32.   Fesik, S. W. (2005) Promoting apoptosis as a strategy for cancer drug discovery, *Nat. Rev. Cancer 5*, 876-885.

33.   Hayflick, L. (1974) Strategy of senescence, *Gerontologist 14*, 37-45.

34.   Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S., and Wright, W. E. (1998) Extension of life-span by introduction of telomerase into normal human cells, *Science 279*, 349-352.

35.   Guittat, L., Alberti, P., Gomez, D., De Cian, A., Pennarun, G., Lemarteleur, T., Belmokhtar, C., Paterski, R., Morjani, H., Trentesaux, C., Mandine, E., Boussin, F., Mailliet, P., Lacroix, L., Riou, J. F., and Mergny, J. L. (2004) Targeting human telomerase for cancer therapeutics, *Cytotechnology 45*, 75-90.

36.   Baluk, P., Hashizume, H., and McDonald, D. M. (2005) Cellular abnormalities of blood vessels as targets in cancer, *Curr. Opin. Genet. Dev. 15*, 102-111.

37.   Nagy, J. A., Chang, S. H., Shih, S. C., Dvorak, A. M., and Dvorak, H. F. (2010) Heterogeneity of the Tumor Vasculature, *Semin. Thromb. Hemost. 36*, 321-331.

38.   Baguley, B. C. (2003) Antivascular therapy of cancer: DMXAA, *Lancet Oncol. 4*, 141-148.

39.   Kerbel, R., and Folkman, J. (2002) Clinical translation of angiogenesis inhibitors, *Nat. Rev. Cancer 2*, 727-739.

40.   Nguyen, D. X., Bos, P. D., and Massague, J. (2009) Metastasis: from dissemination to organ-specific colonization, *Nat. Rev. Cancer 9*, 274-U265.

41.   Berdasco, M., and Esteller, M. (2010) Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry, *Dev. Cell 19*, 698-711.

42.   Jones, P. A., and Baylin, S. B. (2007) The epigenomics of cancer, *Cell 128*, 683-692.

43.   Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps, *Nat. Rev. Genet. 8*, 286-298.

44.   Vogelstein, B., and Kinzler, K. W. (2004) Cancer genes and the pathways they control, *Nat. Med. 10*, 789-799.

45.   Davar, D., Beumer, J. H., Hamieh, L., and Tawbi, H. (2012) Role of PARP Inhibitors in Cancer Biology and Therapy, *Curr. Med. Chem. 19*, 3907-3921.

46.   Drew, Y., and Plummer, R. (2009) PARP inhibitors in cancer therapy: Two modes of attack on the cancer cell widening the clinical applications, *Drug Resist. Update 12*, 153-156.

47.   Jagtap, P., and Szabo, C. (2005) Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors, *Nat. Rev. Drug Discov. 4*, 421-440.

48.   DeNardo, D., Andreu, P., and Coussens, L. M. (2010) Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity, *Cancer Metastasis Rev. 29*, 309-316.

49.   Grivennikov, S. I., Greten, F. R., and Karin, M. (2010) Immunity, Inflammation, and Cancer, *Cell 140*, 883-899.

50. Fiaschi, T., and Chiarugi, P. (2012) Oxidative stress, tumor microenvironment, and metabolic reprogramming: a diabolic liaison, *International journal of cell biology 2012*, 762825.

51. Warburg, O. (1956) Origin of cancer cells, *Science 123*, 309-314.

52. Wallace, D. C. (2012) Mitochondria and cancer, *Nat. Rev. Cancer 12*, 685-698.

53. Aft, R. L., Zhang, F. W., and Gius, D. (2002) Evaluation of 2-deoxy-D-glucose as a chemotherapeutic agent: mechanism of cell death, *Br. J. Cancer 87*, 805-812.

54. Vajdic, C. M., and van Leeuwen, M. T. (2009) Cancer incidence and risk factors after solid organ transplantation, *International Journal of Cancer 125*, 1747-1754.

55. Blattman, J. N., and Greenberg, P. D. (2004) Cancer immunotherapy: A treatment for the masses, *Science 305*, 200-205.

56. Sharma, S. V., and Settleman, J. (2010) Exploiting the balance between life and death: Targeted cancer therapy and "oncogenic shock", *Biochem. Pharmacol. 80*, 666-673.

57. Dougan, M., and Dranoff, G. (2009) Immune Therapy for Cancer, In *Annual Review of Immunology*, pp 83-117, Annual Reviews, Palo Alto.

58. Das, M., Mohanty, C., and Sahoo, S. K. (2009) Ligand-based targeted therapy for cancer tissue, *Expert Opin. Drug Deliv. 6*, 285-304.

59. Tzakos, A. G., Briasoulis, E., Thalhammer, T., Jager, W., and Apostolopoulos, V. (2013) Novel oncology therapeutics: targeted drug delivery for cancer, *Journal of drug delivery 2013*, 918304.

60. Green, M. R. (2004) Targeting targeted therapy, *N. Engl. J. Med. 350*, 2191-2193.

61. Munagala, R., Aqil, F., and Gupta, R. C. (2011) Promising molecular targeted therapies in breast cancer, *Indian J. Pharmacol. 43*, 236-245.

62. Hemminki, A. (2002) From molecular changes to customised therapy, *Eur. J. Cancer 38*, 333-338.

63. Hoelder, S., Clarke, P. A., and Workman, P. (2012) Discovery of small molecule cancer drugs: Successes, challenges and opportunities, *Mol. Oncol. 6*, 155-176.

64. Sikora, K. (2001) Cancer drug development in the post-genomic age, *Curr. Sci. 81*, 549-554.

65. Chari, R. V. J. (2008) Targeted cancer therapy: Conferring specificity to cytotoxic drugs, *Accounts of Chemical Research 41*, 98-107.

66.   Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) The refined 2.15-A X-ray crystal-structure of human liver Cathepsin-B - The structural basis for its specificity, *Embo J. 10*, 2321-2330.

67.   Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997) Role of the occluding loop in cathepsin B activity, *J. Biol. Chem. 272*, 1197-1202.

68.   Schenker, P., Alfarano, P., Kolb, P., Caflisch, A., and Baici, A. (2008) A double-headed cathepsin B inhibitor devoid of warhead, *Protein Sci. 17*, 2145-2155.

69.   Song, J. X., Xu, P., Xiang, H., Su, Z. D., Storer, A. C., and Ni, F. (2000) The active-site residue Cys-29 is responsible for the neutral-pH inactivation and the refolding barrier of human cathepsin B, *FEBS Lett. 475*, 157-162.

70.   Turk, B., Turk, D., and Salvesen, G. S. (2002) Regulating cysteine protease activity: Essential role of protease inhibitors as guardians and regulators, *Curr. Pharm. Design 8*, 1623-1637.

71.   Gondi, C. S., and Rao, J. S. (2013) Cathepsin B as a cancer target, *Expert Opin. Ther. Targets 17*, 281-291.

72.   Joyce, J. A., and Hanahan, D. (2004) Multiple roles for cysteine cathepsins in cancer, *Cell Cycle 3*, 1516-1519.

73.   Nomura, T., and Katunuma, N. (2005) Involvement of cathepsins in the invasion, metastasis and proliferation of cancer cells, *The journal of medical investigation : JMI 52*, 1-9.

74.   Mort, J. S., and Buttle, D. J. (1997) Cathepsin B, *Int J. Biochem. Cell Biol. 29*, 715-720.

75.   Bagatell, R., and Whitesell, L. (2004) Altered Hsp90 function in cancer: A unique therapeutic opportunity, *Mol. Cancer Ther. 3*, 1021-1030.

76.   Kim, L. S., and Kim, J. H. (2011) Heat Shock Protein as Molecular Targets for Breast Cancer Therapeutics, *J. Breast Canc. 14*, 167-174.

77.   Wang, J. H., Cui, S. Z., Zhang, X. L., Wu, Y. B., and Tang, H. S. (2013) High Expression of Heat Shock Protein 90 Is Associated with Tumor Aggressiveness and Poor Prognosis in Patients with Advanced Gastric Cancer, *PLoS One 8*.

78.   Li, J., and Buchner, J. (2013) Structure, function and regulation of the hsp90 machinery, *Biomedical journal 36*, 106-117.

79.   Prodromou, C. (2012) The 'active life' of Hsp90 complexes, *Biochim. Biophys. Acta-Mol. Cell Res. 1823*, 614-623.

80.   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.

81.   Wang, H., Tan, M. S., Lu, R. C., Yu, J. T., and Tan, L. (2014) Heat Shock Proteins at the Crossroads between Cancer and Alzheimer's Disease, *Biomed Res. Int.*

82.   Didenko, T., Duarte, A. M. S., Karagoz, G. E., and Rudiger, S. G. D. (2012) Hsp90 structure and function studied by NMR spectroscopy, *Biochim. Biophys. Acta-Mol. Cell Res. 1823*, 636-647.

83.   Zhang, L., Fok, J. H. L., and Davies, F. E. (2014) Heat shock proteins in multiple myeloma, *Oncotarget 5*, 1132-1148.

84.   Murray, C. W., Carr, M. G., Callaghan, O., Chessari, G., Congreve, M., Cowan, S., Coyle, J. E., Downham, R., Figueroa, E., Frederickson, M., Graham, B., McMenamin, R., O'Brien, M. A., Patel, S., Phillips, T. R., Williams, G., Woodhead, A. J., and Woolford, A. J. A. (2010) Fragment-Based Drug Discovery Applied to Hsp90. Discovery of Two Lead Series with High Ligand Efficiency, *Journal of Medicinal Chemistry 53*, 5942-5955.

85.   Isaacs, J. S., Xu, W. P., and Neckers, L. (2003) Heat shock protein 90 as a molecular target for cancer therapeutics, *Cancer Cell 3*, 213-217.

86.   Powers, M. V., and Workman, P. (2007) Inhibitors of the heat shock response: Biology and pharmacology, *FEBS Lett. 581*, 3758-3769.

87.   Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis, *Nature 475*, 324-332.

CHAPTER 3

3. Introduction to computational chemistry

This chapter reviews the computational chemistry methods used in this study. It also provide with some theoretical background pertaining to computational chemistry and computer-aided drug design (CADD).

3.1. Introduction

Drug discovery and development is a very costly and a time-consuming process. The estimated time frame for drug discovery, from the early stages of identification to clinical trials is 10-15 years (*1*). One of the main problems in computational chemistry is late stage drug failure, which is mainly due to absorption, distribution, metabolism, excretion and toxicity (ADME/Tox) (*2*). Computer-aided drug design was introduced in the 1970s to help minimize drug failure problems, and accelerates the drug discovery process. CADD exploits computational tools and resources that are suitable for compound analysis, storage, management, and modelling (*2-4*).

The main methods describing computational chemistry are quantum mechanics (QM) and molecular dynamics (MD). Quantum mechanics is a division of mechanics that explains the motion and interaction of subatomic particles using mathematical descriptions (*5, 6*). Molecular dynamics is a computer simulation, where atoms and molecules are allowed to interact for a period of time, thereby providing a view of their physical motion in the context of N-body simulation. One aspect of MD is molecular mechanics (MM), which is based on the use of classical or Newtonian mechanics to describe the physical basis behind the models (a model of a molecule is referred to as a collection of atoms held together by bonds) (*7*). In this thesis, MM and MD were used to predict conformational changes that may occur in the biological target when the ligand is bound to it. In the next sections, computational theories in MM and MD are described, namely, the Born-Oppenheimer, Schrodinger's equation, potential energy surface (PES), and MD methodologies applied in this research study.

3.2. The Schrödinger equation

The Schrödinger equation is an analogue of the 2nd law of motion in classical mechanics, and describes the motion and behavior of systems on the atomic and subatomic levels using the wave

function (*2*). The time dependent Schrödinger equation is a widely used equation that contends that the wavefunction does not only depends on position but also on time (*8, 9*). The simplest form of the Schrödinger equation is denoted as:

$$H\psi = E\psi \qquad \text{Eq. 1}$$

Where H is the molecular Hamiltonian, ψ a wavefunction (describes the state of a complete system) and E is the energy of the system (*10, 11*). The molecular Hamiltonian is a sum of the kinetic (T) and potential (V) energy, as shown below:

$$H = T + V \qquad \text{Eq. 2}$$

The Hamiltonian is presented in detail as:

$$H = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{\hbar^2}{2}\sum_A \frac{1}{M_A}\nabla_A^2 - \sum_A \sum_i \frac{Z_A e^2}{r_{Ai}} + \sum_i \sum_{j>i}\frac{e^2}{r_{ij}} + \sum_A \sum_{B>A}\frac{Z_A Z_B e^2}{R_{AB}} \qquad \text{Eq. 3}$$

where i; j refer to electrons and A;B refer to nuclei.

3.3. Born-Oppenheimer approximation

The Born-Oppenheimer approximation was proposed by M. Born and J. R. Oppenheimer in 1927 (*12*), and it simplifies the complexity of the Schrödinger equation. The nucleus exerts a larger mass in comparison to the electrons, thus the nuclei moves much slowly then the electrons. The Born-Oppenheimer approximation takes advantage of this phenomenon and assumes the nucleus to be stationary while electrons move around it (*8, 9*). Therefore, the approximation states that the Schrödinger equation (section 3.2) may be separated into an electronic and nuclear equation by solving the wavefunction (*12*), which then becomes:

$$\psi(r_{molec}) = \psi(r_{elec})\,(r_{nucl}) \qquad \text{Eq. 4}$$

Eq. 1 is then presented as:

$$H_{EN}\psi(r_{el}) = E_{EN}\psi(r_{el}) \qquad \text{Eq. 5}$$

In H_{EN} approximation, there is a difference between the terms due to their reliance on fixed nuclear position (V_{NN}) or the reliance on the not fixed electron position (H_{el}). Eq. 5 includes the energy term E_{EN} derived from the source that is variable due to the changing electron coordinates (*E*), and the one that is constant due to fixed nuclear coordinates (*N*):

$$(H_{el} + V_{NN})\psi(r_{el}) = E_{EN}\psi(r_{el}) \qquad \text{Eq. 6}$$

The Born-Oppenheimer approximation is better for ground state molecules than it is for the excited state. From the approximated wavefunction and Hamiltonian, it is possible to obtain the energy of the molecule, normally called the "electron energy including nuclear repulsion", which can be obtained by constructing the molecular energy curve (*13*). The direct result of the Born-Oppenheimer approximation is the potential energy surface (PES).

3.4. Potential energy surface

Potential energy surface (PES) refers to the mathematical or graphical relationship between the energy of a molecule and its geometry (*14*). The PES concept originates from the Born-Oppenheimer approximation, which states that the nuclei in a molecule are stationary compared to the electrons, thus making a molecular shape or geometry concept significant (*8, 14, 15*). Therefore, PES aids visualization and understanding of the potential energy and molecular geometry relationship, and provides an insight on how computational chemistry programs, locate and characterize structures of interest (*8*). The difference in the nuclei and electron motion leads to the so-called electronic adiabatic potential energy surfaces (*14*). Moreover, PES is merely a plot of the energy of the nuclei and electrons against the geometric coordinates of the nuclei (nuclei position is mostly the ground-state energy surface) (**Figure 3.1**) (*8, 11*).
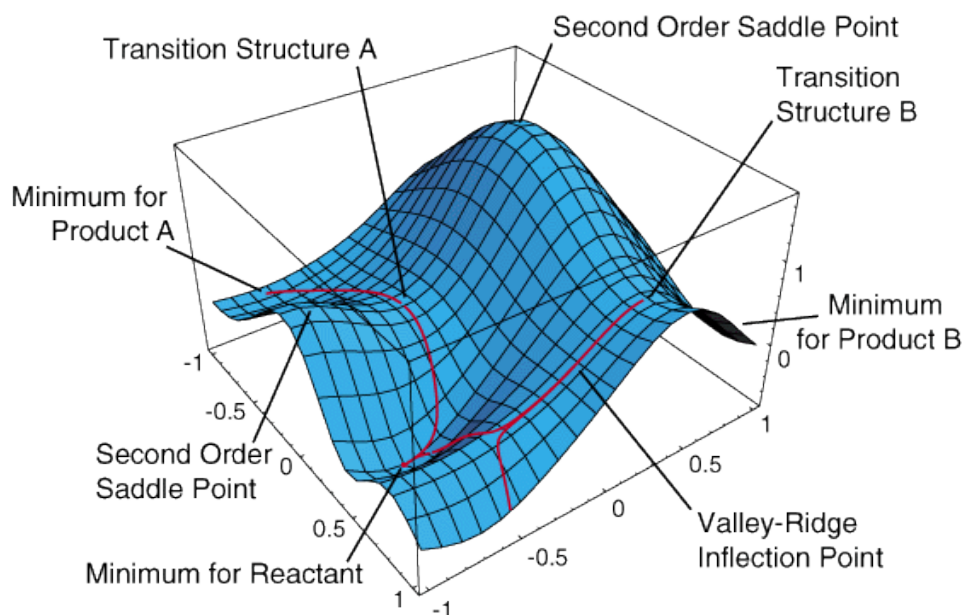
Figure 3.1: Graphical representation of a two-dimensional potential energy surface (*16*).

3.5. Molecular mechanics

Molecular mechanics is a method used to determine geometry, enthalpies, vibrational spectra and molecular energies of formations of stable ground state molecules (*8, 9*). Molecular mechanics is mainly based on classical models to predict the energy of a molecule as a function of its conformation, providing insights into the changes in cellular structure, response and function. As a result, this understanding offers new opportunities for disease diagnosis and treatment (*8*). This method is widely used for computing the energies and geometries of large biological molecules, such as proteins and nucleic acids, in which approximately 100 000 atoms can be treated and simulations over time scales of tens of nanoseconds can be achieved (*8, 17*). The total potential energy of a molecule is described as the sum of various interactions, namely bonded interactions, such as bond angles, bond length, torsion and non-bonded interactions, such as van der Waals and electrostatic interactions, as described in the following equation:

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdw} + E_{elec}$$ Eq. **7**

Here E_{tot} is the total potential energy, E_{str} the bond-stretching energy, E_{bend} the bond angle-bending energy, E_{tor} the torsional or twisting energy, E_{vdw} is the van der Waals forces and E_{elec} is the electrostatic forces between the non-bonded atoms.

3.6. Force Field

A force field refers to the combination of a mathematical equation and the associated parameters used to describe the dependence of the energy of a system on its atomic coordinates. The concept of force fields first appeared in 1960's with the development of MM methods (*18, 19*). It entails functional form of the interatomic potential energy and a parameter set that enter into this form. These parameters are either obtained from semi-empirical quantum mechanical calculations or from experimental data such as Raman, NMR, X-ray, neutron spectroscopy and electron diffraction (*18*). The widely used functional form and parameterization protocols of biomolecular force fields involve AMBER, CHARM, NAMD, GROMOS and OPLS-AA. In the current study, AMBER force field was used, applying the General AMBER Force Field (GAFF) parameters for the ligand and standard AMBER force field for the enzyme (*20, 21*).

3.7. Molecular Dynamics

Due to advances in computer technology and algorithmic improvements, MD has become a valuable technique for computer simulating complex systems and assessing the thermodynamics of molecules (*22, 23*). Molecular dynamics simulations provide a deeper understanding of the nature of ligand-protein interactions, and present such interactions as ensembles that follow the Law of Statistical Dynamics (*22, 24*). MD simulations produce a dynamical trajectory for a system by numerically solving the Newton's Laws of motion for the interacting species (*18, 25*). The results of the MD simulations are characterized as a trajectory that specifies particles position and velocity at different periods of time (*26*). Therefore, the Force (F_i) operating on a particle is determined as a function of time that equals to the negative gradient of potential energy:

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \qquad \text{Eq. 8}$$

Here $r_i(t) = (x_i(t), y_i(t), z_i(t))$ is the (i) particle position vector, F_i the force acting on (i) particle at time t and m_i the mass of a particle. According to the Newton's Law of Motion, acceleration (α) of a particle is obtained from the force divided by a mass of a particle, as depicted in the following equation:

$$a = -\frac{F_i}{m_i} \qquad \text{Eq. 9}$$

The change in the velocity of a particle is equal to the integral of acceleration over a period of time, and the position change is equal to the integral of velocity over a period of time, as shown in the following equation:

$$dv = \int a \, dt, \qquad \qquad \text{Eq. 10}$$

$$dr = \int v \, dt \qquad \qquad \text{Eq. 11}$$

As a result, the kinetic energy can be defined in terms of the velocity and momenta of the given particles:

$$K(v) = \frac{1}{2}\sum_{i=1}^{N} m_i v_i \qquad \qquad \text{Eq. 12}$$

$$K(p) = \frac{1}{2}\sum_{i=1}^{N} \frac{p^2{}_i}{m_i} \qquad \qquad \text{Eq. 13}$$

Therefore, the total energy of a given system is given as Hamiltonian (H), the sum of kinetic and potential energies, as represented in the equation below:

$$H(q,p) = K(p) + U(p) \qquad \qquad \text{Eq. 14}$$

Where, q is a set of Cartesian coordinates, p denotes the momenta of the given particles and $U(q)$ is the potential energy function. The first derivatives of the positions with respect to time are the velocities and can be determined using the equation represented below:

$$v_i(t) = \frac{d}{dt} q_i(t) \qquad \qquad \text{Eq. 15}$$

Where $qi(t)$ 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 (*25, 26*).

3.8. Approaches for estimating binding affinities

3.8.1. Molecular Docking

Molecular docking is a computational tool that is essential for structure-based drug design (SBDD). These programs are commonly used to predict the binding modes and affinities of

small molecules (ligands) interacting with the binding site of the target protein of the known 3D structure (*27-29*). Molecular docking can be applied in binding-site identification (blind docking), protein-protein interaction, enzymatic reaction mechanisms, drug discovery (lead optimization), protein engineering, virtual screening (hit identification) and de-orphaning of the receptor. Docking is achieved through two correlated steps: firstly, by sampling conformations of the ligands in the active site of the protein, secondly, by ranking the conformations via a scoring function. It is highly recommended that the scoring function should rank the highest among all generated conformations and that the sampling algorithms should be able to reproduce the experimental data (*30, 31*). Thus, docking can be used to screen large libraries of compounds, rank the results, and then formulate the structural hypothesis of the ligand-receptor inhibitory mode, which is essential in lead optimization. The ligand-receptor binding energy is analyzed using molecular mechanics presented in the following equation:

$$E_{binding} = E_{target} + E_{ligand} + E_{target-ligand} \qquad \text{Eq. 16}$$

Some examples of molecular docking programs that are normally used for academic and commercial purposes include: Dock (*27*), AutoDock (*32*), GOLD (*28*), GLIDE (*33*), Surflex (*34*), ICM (*35*) and PhDOCK (*34*). Although these programs entail different information regarding the structures of component proteins, they have common computational steps, namely simplified and/or rigid body search, selecting the region(s) of interest, refinement of docked structures and selecting the best models (*35*). In the current studies, a quicker version of AutoDock and AutoDock vina (*36*) was used. This allowed the prediction of flexible ligands' binding affinity to rigid Cathepsin B and HSP90 active sites (refer to chapter 4 and 5). In chapter 4, we also used a CovalentDock cloud (*37*) to screen large libraries of compounds that form covalent bond with the target protein Cathepsin B.

3.8.2. Binding free energy calculations

Free binding energy calculations have been proven useful in the computational biochemical research, as they enable detailed investigation of binding affinities of small drug-candidate compounds. They also allow an evaluation of relative stabilities of large biomacromolecular structures (*38, 39*). The ligand-receptor binding affinities prediction serves as a cornerstone in the study of protein-protein interaction to structure-based drug design (*40*). The Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics

Generalized Born Surface Area (MM/GBSA) are efficient methods for calculating free energies of numerous molecular systems (*41*). In the current studies, MM/GBSA calculations were applied for confirmation of the binding energies. A brief description about free binding energy calculations is provided in chapter 5.

3.8.3. Entropy calculations

Entropy is defined as the measure of dynamics of the overall system and it forms important aspects of thermodynamics properties that play an essential role in the enzyme catalytic effect (*22*). The entropic contributions to the free energy reflect the strength and specificity of interactions between the enzyme-ligand complexes. Moreover, alterations in the binding entropy reveal a loss of motion due to changes in rotational and translational degrees of freedom of the interacting molecules (*42, 43*). Entropic effects have been shown to enable enzymes to speed up their reaction, thus these findings were essential in understanding the origins of enzyme catalysis (*42*). A recent study has shown that the entropies of activation in the enzyme and in water are related, with the overall catalysis being mainly due to enthalpic effects (*44*). The Gibb's equation indicates that entropy can be calculated directly from ΔG and ΔH. However, the entropy associated with ligand-protein binding can be expressed as the sum of numerous contributing effects, at a given temperature (*22*). The entropy is best described by the following equation:

$$\Delta G = \Delta H - T\Delta S \qquad \text{Eq. 17}$$

Therefore,

$$\Delta S = \frac{\Delta G}{\Delta S - T} \qquad \text{Eq. 18}$$

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

3.9. Molecular Modelling methods used in this study

3.9.1. Homology Modelling

A deep understanding of the conformation of multi-domain proteins, DNA-protein complexes and large multi-protein complexes is important to explore various biological problems. To achieve this, nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography and electron microscopy are used to predict the 3D structure of macromolecules. However, such

techniques require expert-user knowledge in order to obtain accurate results (*45, 46*). Therefore, homology modelling has been considered the easier technique for predicting the 3D protein structure when only sequence data is available. Homology modelling builds the protein structure using the sequence of a protein whose X-ray structure is known as a building template. The success of the homology modelling is determined by the following factors: the ability to detect the existing homologue structure; the existence of a known homologue with known structure; the ability to detect the homologue and the quality of the model building.

Programs available for homology modelling utilize different methods to produce a final model. Within each step of the homology modelling process, many factors affect the quality of the model produced and appropriate selection of the program can significantly improve the quality of the model (*47*). During the selection of homologues in homology modelling, the crystallographic-defined structures of protein are selected, with a resolution ranging from 1-3Å. In addition, High-resolution structures (with resolution values of 1Å) are highly ordered, and it is easy to see every atom in the electron density map, while lower resolution structures (with resolution of 3Å or higher) show only the basic outlines of the protein chain with inferred view of the atomic structure. In this research study (chapter 5), the homology model of human Hsp90 was constructed following the steps outlined in Figure 3.2 below.

```
┌─────────────────────────┐
│ Download protein crystal │
│   structure from PDB    │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Query protein sequence and │
│ UniProt ID from Swiss model │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ BLAST protein sequence to │
│ query protein homologues │
└─────────────────────────┘
            │
            ▼
      ╭─────────────╮
      │   Protein   │
      │  homologue  │
      │ template(s) >1 │
      ╰─────────────╯
```

┌──────────────────────────┐   │
│ Select suitable homologue of │   ▼
│ good quality between 1-3Å │   ┌─────────────────────┐
│   resolution   │   │ Align and superimpose │
└──────────────────────────┘   │ protein PDB on chimera │
                               └─────────────────────┘
                                        │
                                        ▼
                               ┌─────────────────────┐
                               │   Open main protein   │
                               │   sequence   │
                               └─────────────────────┘
                                        │
                                        ▼
                               ┌─────────────────────┐
                               │ Align the PDB sequence │
                               │   and main sequence   │
                               └─────────────────────┘
                                        │
                                        ▼
                               ┌─────────────────────┐
                               │   **Build model**   │
                               └─────────────────────┘
                                        │
                                        ▼
┌──────────────────┐   ┌─────────────────────┐
│ Save final model │ ◄────── │ Delete undesirable   │
│   as PDB   │   │ sections from the newly │
└──────────────────┘   │ built protein sequence │
                               └─────────────────────┘

Figure 3.2: Schematic representation of protocol followed during construction of Hsp90 homology model.

3.9.2. Per-residue decomposition

Per-residue decomposition was developed to determine the specific energy contribution of each residue (also referred to as "hot spots") towards ligand binding. The "hot spots" idea is that a number of amino acids at a binding interface make a dominant contribution to the binding affinity (*48-50*). However, further understanding of the roles of residues surrounding the "hot spots" is required in order to create a favorable binding environment. Per-residue decomposition is also used to determine bond formation in a molecule-molecule interaction (*51*). In per-residue decomposition, the contributions of rotational and translational entropies, on a residue basis, are expected to be insignificant or constant.

References

1. 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, *Current Topics in Medicinal Chemistry 10*, 127-141.

2. Song, C. M., Lim, S. J., and Tong, J. C. (2009) Recent advances in computer-aided drug design, *Brief. Bioinform. 10*, 579-591.

3. Kore, P. K., Mutha, M. M., Antre, R. V., Oswal, R. J., and Kshirsagar, S. S. (2012) Computer-Aided Drug Design: An Innovative Tool for Modeling, *Medicinal Chemistry 2*, 139-148.

4. Marshall, G. R. (1987) Computer-aided drug design, *Annu. Rev. Pharmacol. Toxicol. 27*, 193-213.

5. Gray, S. (2008) Introduction to quantum mechanics - A time-dependent perspective, *Science 319*, 161-161.

6. Hewitt-Horsman, C. (2009) An Introduction to Many Worlds in Quantum Computation, *Found. Phys. 39*, 869-902.

7. Bartels, C. (2000) Analyzing biased Monte Carlo and molecular dynamics simulations, *Chem. Phys. Lett. 331*, 446-454.

8. Lewars, E. G. (2004) *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics*, 2nd ed., Kluwer Academic Publishers.

9. Rogers, D. W. (2003) *Computational chemistry using PC*, 3rd ed., John Wiley & Sons

10. Jensen, F. (1999) *Introduction to Computational Chemistry*, 2nd ed., Wiley and Sons.

11. Le Bris, C. (2005) Computational chemistry from the perspective of numerical analysis, *Acta Numerica 14*, 363-444.

12. Born, M., and Oppeinheimer, R. (1927) Zur Quantentheorie der Molekeln, *Annalin Physik 84*, 457-484.

13. Atkins, P. (2005) *Molecular quantum mechanics*, 4th ed., Oxford University Press Inc.

14. Truhlar, D. G. (2001) Potential energy surfaces, *The Encyclopaedia of Physical Science and Technology 13*, 9-17.

15. Wales, D. J., and Bogdan, T. V. (2006) Potential energy and free energy landscapes, *Journal of Physical Chemistry B 110*, 20765-20776.

16.   http://www.chem.wayne.edu/~hbs/chm6440/PES.html, 13 Nov. 2014

17.   Senn, H. M., and Thiel, W. (2009) QM/MM Methods for Biomolecular Systems, *Angew. Chem.-Int. Edit. 48*, 1198-1229.

18.   Gonzalez, M. A. (2011) Force fields and molecular dynamics simulations, *Science 12*, 169-200.

19.   Guvench, O., and MacKerell, A. D., Jr. (2008) Comparison of protein force fields for molecular dynamics simulations, In *Methods in Molecular Biology* (Kukol, A., Ed.), pp 63-88, Humana Press Inc, 999 Riverview Dr, Ste 208, Totowa, Nj 07512-1165 USA.

20.   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, *Journal of Computational Chemistry 26*, 1668-1688.

21.   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 progams fo applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energrtic properties of molcules, *Computer Physics Communications 91*, 1-41.

22.   Bronowska, A. K. (2011) Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design, *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*.

23.   Meller, J. (2001) Molecular Dynamics, *Encyclopaedia of Life Sciences*.

24.   McCommon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins, *Nature 267*, 585-590.

25.   Binder, K., Horbach, J., Kob, W., Wolfgang, P., and Varnik, F. (2004) Molecula dynamics simulations, *Journal of Physics: Condensed Matter 16*.

26.   Lipkowitz, K. B., and Donald, D. B. (1997) *Reviews in computational chemistry*, Vol. 18, VCH Publishers, New York.

27.   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.

28.   Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD, *Proteins 52*, 609-623.

29. Vijesh, A. M., Isloor, A. M., Telkar, S., Arulmoli, T., and Fun, H. K. (2013) Molecular docking studies of some new imidazole derivatives for antimicrobial properties, *Arab. J. Chem. 6*, 197-204.

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. Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011) Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery, *Current Computer-Aided Drug Design 7*, 146-157.

32. 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.

33. 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, *Journal of Medicinal Chemistry 47*, 1739-1749.

34. Joseph-McCarthy, D., and Alvarez, J. C. (2003) Automated generation of MCSS-derived pharmacophoric DOCK site points for searching multiconformation databases, *Proteins 51*, 189-202.

35. Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y. B., and Humblet, C. (2009) Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy, *Journal of Chemical Information and Modeling 49*, 1455-1474.

36. 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, *Journal of Computational Chemistry 19*, 1639-1662.

37. Ouyang, X. C., Zhou, S., Su, C. T. T., Ge, Z. M., Li, R. T., and Kwoh, C. K. (2013) CovalentDock: Automated covalent docking with parameterized covalent linkage energy

estimation and molecular geometry constraints, *Journal of Computational Chemistry 34*, 326-336.

38.   Homeyer, N., and Gohlke, H. (2012) Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method, *Mol. Inf. 31*, 114-122.

39.   Miller, B. R., McGee, T. D., 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.

40.   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, *Biophysical Journal 86*, 67-74.

41.   Sun, H. Y., Li, Y. Y., Tian, S., Xu, L., and Hou, T. J. (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set, *Phys. Chem. Chem. Phys. 16*, 16719-16729.

42.   Page, M. I. (1977) Entropy, binding-energy, and enzymic catalysis, *Angew. Chem.-Int. Edit. Engl. 16*, 449-459.

43.   Villa, J., Strajbl, 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.

44.   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.

45.   Dalton, J. A. R., and Jackson, R. M. (2010) Homology-Modelling Protein-Ligand Interactions: Allowing for Ligand-Induced Conformational Change, *Journal of Molecular Biology 399*, 645-661.

46.   Nollmann, M., Stark, W. M., and Byron, O. (2005) A global multi-technique approach to study low-resolution solution structures, *J. Appl. Crystallogr. 38*, 874-887.

47.   Pitman, M. R., and Menz, R. I. (2006) Methods for Protein Homology Modelling, In *Applied Mycology and Biotechnology, Vol 6: Bioinformatics* (Arora, D. K., Berka, R. M.,

and Singh, G. B., Eds.), pp 37-59, Elsevier Science Bv, Sara Burgerhartstraat 25, Po Box 211, 1000 Ae Amsterdam, Netherlands.

48.   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, *Journal of Molecular Biology 330*, 891-913.

49.   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-Structure Function and Bioinformatics 67*, 418-434.

50.   Zoete, V., Meuwly, M., and Karplus, M. (2005) Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition, *Proteins-Structure Function and Bioinformatics 61*, 79-93.

51.   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-Structure Function and Bioinformatics 67*, 1026-1047.

CHAPTER 4

In-silico identification of irreversible cathepsin B inhibitors as anti-cancer agents: virtual screening, covalent docking Analysis and molecular dynamics simulations.

Mbatha Sbongile,[a] and Mahmoud E. S. Soliman[a*]

[a]Discipline 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

Cathepsin B is a cysteine protease that belongs to the papain superfamily. Malfunctions related to cathepsin B can lead to inflammation and cancer. Via an integrated *in-silico* approach, this study is aimed to identify novel Michael acceptors-type compounds that can irreversibly inhibit cathepsin B enzyme via covalent bond formation with the active site cysteine residue. Here, we report the first account of covalent docking approach incorporated into a hybrid ligand/structure-based virtual screening to estimate the binding affinities of various compounds from chemical databases against the cathepsin B protein. For validation, compounds with experimentally determined anti-cathepsin B activity from PubChem bioassay database were also screened and covalently docked to the enzyme target. Interestingly, four novel compounds exhibited better covalent binding affinity when compared against the experimentally determined prototypes. Molecular dynamics simulations were performed to ensure the stability of the docked complexes and to allow further analysis on the MD average structures. Per-residue interaction decomposition analysis was carried out to provide deeper insight into the interaction themes of discovered hits with the active site residues. It is found that polar and hydrophobic interactions contributed the most towards drug binding.

The hybrid computational methods applied in this study should serve as a powerful tool in the drug design and development process.

Key Words:

Cathepsin B; Michael acceptors; virtual screening; covalent docking; molecular dynamics

Running tittle:

Identification of Michael acceptors as irreversible cathepsin B inhibitors.

1. Introduction

Cancer is considered as one of the leading cause of deaths worldwide. The American Cancer Society reported that cancer constitutes about 2-3% of deaths globally (*1*). It has been reported that anti-cancer drugs showed undesirable side effect on chronic therapy in patients (*1*). Therefore, there is a constant need to develop new anti-cancer drugs with higher therapeutic index and improved patients compliance.

Cathepsins are lysosomal peptidases responsible for intracellular proteolysis in mammalian cells (*2*). These enzymes are dependent on a nucleophilic thiol group (-SH) of a cysteine residue for their enzymatic activity (*3*) (see figure 1a). They also have been found to be involved in biological processes related to variant diseases in humans (*4*). One of the lysosomal cysteine proteases is known as Cathepsin B and it belongs to the papain superfamily (*5*). Cathepsin B has been reported to be involved in human diseases such as cardiovascular disease, neurodegenerative disorder, inflammation and cancer. These diseases are due to the alterations in cathepsin B expression, activity, localization and protein levels (*6*). This enzyme has the ability to facilitate cell migration leading to tumor metastasis and angiogenesis resulting to cancer progression (*2, 4, 7*). As a result, design of cathepsin B inhibitors is highly considered in the development of therapeutic agents (*2, 5*).

Biological studies have shown that irreversible binding occurs between the enzyme and several cathepsin inhibitors (*4*). Most of the cathepsins contain an active site cysteine residue (Cys29) that acts as a nucleophile and a histidine residue (His199) that acts as a general base in the hydrolysis of a target peptide bond in the substrate (*6, 8*). Compounds that contain Michael acceptors are considered as one of the most active irreversible cysteine protease inhibitors (*9*).

These inhibitors include α, β unsaturated carbonyl derivatives, vinyl sulfones and various derivatives (Figure 1a). Such inhibitors bind to the active site of the enzyme forming a covalent bond with the catalytic thiol group of the enzyme (*4*).

This study is designed to further identify novel Michael acceptor-type leads from commercial chemical database as well as to investigate their relative binding affinities and themes with cathepsin B by a robust integrated *in-silico* approach.

Conventional virtual screening using non-covalent docking is widely used in drug discovery research (*10-13*) including the search for cathepsin B inhibitors (*2, 7*). However, in this study, we adopted a novel and time-efficient hybrid non-covalent/covalent docking approach supported by molecular dynamics simulations and post-dynamics per-residue interaction energies calculations in order to provide more precise description of the binding modes of covalent inhibitors. To our knowledge, this is the first study that embarked on a hybrid non-covalent and covalent molecular docking methodology in the virtual screening for new drug leads.

In search for novel inhibitors, ZINC database (http://zinc.docking.org) was used. Despite the fact that the covalent docking approach has been validated on a wide range of biological systems (*14*), we opted to further assess the computational methods and results obtained from our study on a set of similar and experimentally determined molecule with known cathepsin B inhibitory activity (prototypes) obtained from PubChem bioassay database (http://pubchem.ncbi.nlm.nih.gov/pcassay) (see Methods section). Results from the virtual screening were compared against the experimentally determined prototypes.
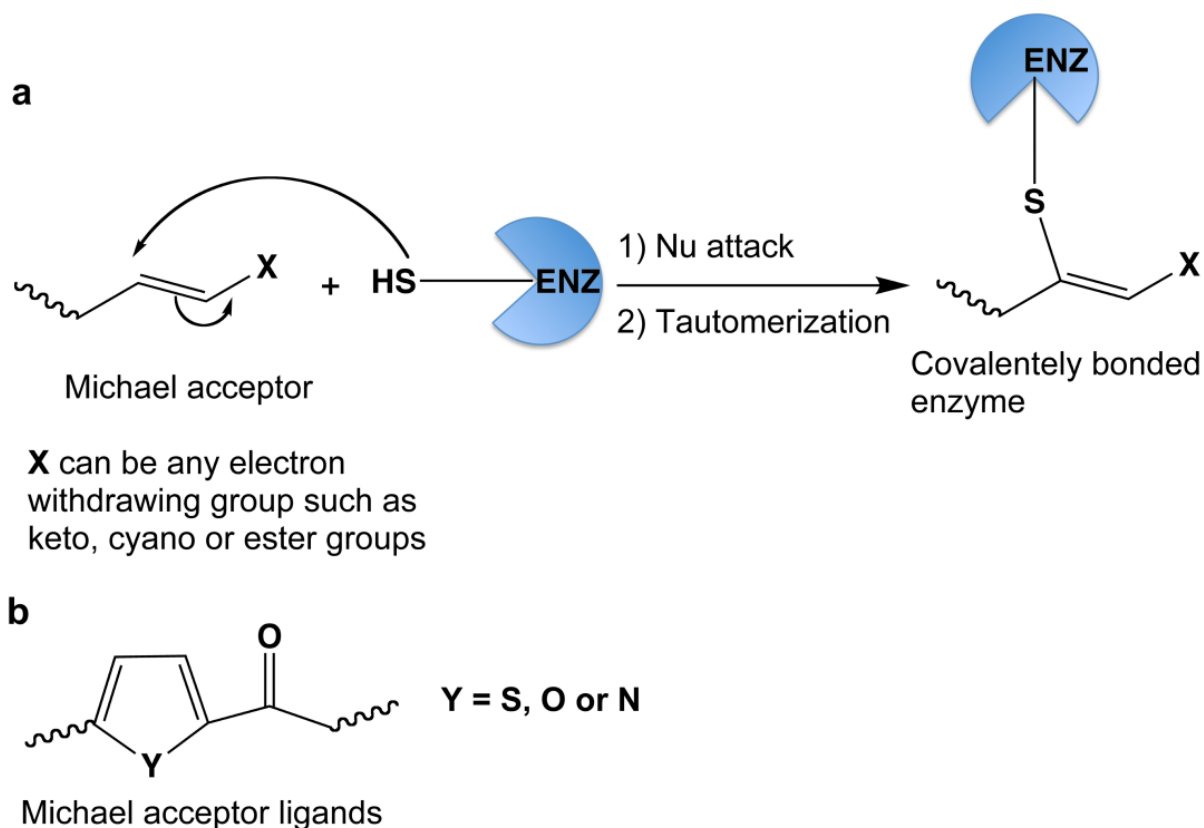
Figure 1: (a) General mechanism of Michael addition reaction with nucleophilic groups in enzymes (such as cysteine). (b) Michael acceptor ligands used in this study.

2. Computational Methodology

The computational procedure adopted in this work is depicted in Figure 2 as well as explained in the text below.

**Target protein
(PDB: 1GMY)**

Physicochemical filters
- Drug-like
- Lipinski's rule

**Mining chemical database
for potential hits based on
a prototype (Michael
acceptor-like)**

- Retrieved compounds from chemical
database were energy minimized then
aligned prior docking
- Structure-based Virtual screening (SBVS)
by docking all ligands against target
protein

**Non-covalent docking
(Autodock Vina)**

- Ligand candidates were ranked
according to binding affinity
- Top-50 ligands were filtered

Ranking & selection

**Top-50 SBVS-generated
hits**

Receptor nucleophilic center
coordinates are defined (active
site Cysteine)

**Covalent docking
(CovalentDock)**

**Ranking & selection
(Top-10)**

Final hits

**Molecular dynamics simulations
and post-dynamic analysis**

Figure 2: Workflow of the computational procedure adopted in this study.

2.1.   Enzyme acquisition

The 3D complex structure of cathepsin B with its inhibitor dipeptidyl nitrile (DPN) was obtained from the protein data bank, PDB code: 1GMY (*2*). The enzyme was then isolated from the ligand using Molegro Molecular Viewer (MMV) software suite (www.nlcbio.com) by deleting subunit C from the enzyme trimer. The free cathepsin B dimer was then used in all subsequent computational simulations. In a previous study, it was reported that simulation of a dimeric form of this enzyme is important as residues from both the first and the second protein chains contributed to the binding site of dipeptidyl nitrile (*2*).

2.2.   Ligands library generation

Our search for potential inhibitors was based on electrophilic Michael acceptor-like compounds with known biological activity as irreversible inhibitors (prototypes) (*2*). It was found that most of these inhibitors have a thiophene-2-one or pyrol-2-one fragment in their chemical structure (*2*) (Figure 1b), therefore, we opted to use this fragment as a basis for pharmacophoric search.

The ZINC database (http://zinc.docking.org) was used for screening of libraries of compounds. MarvinSketch Software (Chemaxon, http://www.chemaxon.com) was used to draw the template structures for the ZINC database queries based on Michael acceptor-like compounds as shown in Figure 1b. The ZINC database was queried for compounds with drug-like properties, which includes the molecular weight ranging from 300 to 500 g/mol, logP value between 1 and 5, maximum of 5 hydrogen bond donor atoms and maximum of 10 bond acceptor atoms. This query gave rise to 1186 compounds hits that were downloaded as a single mol2 file. Molegro

Molecular Viewer software (www.nlcbio.com) was then used to export the single mol2 file of downloaded compounds into separate mol2 files for each compound. These files were then converted into the pdbqt format required for docking using Raccoon software (*15*).

2.3. Structure-based virtual screening and validation of docking protocol

2.3.1. Non-covalent docking

It is worth mentioning that covalent docking is more computationally exhausting and time consuming than conventional non-covalent docking due to the additional energy terms and processing steps in case of the former. Therefore, we opted to run initial non-covalent docking to rank the compounds library from the chemical database based on their relative binding affinities. To this end, all 1186 retrieved compounds, were docked into the active site of cathepsin B using Autodock Vina software (*16*). The grid box was defined using Chimera software (*17*) with the grid parameters being X= 34, Y= 42 and Z= 48 for the dimensions and X= 33.9, Y= 34.1 and Z= 32.9 for the center grid box. The docked compounds were then ranked according to the binding affinities against the target protein. The top-50 ranked compounds resulted from non-covalent docking were then selected for further covalent docking.

2.3.2. Covalent docking

The theory behind covalent docking is fully explained elsewhere (*14*). CovalentDock software (*14*) was used to dock the top-ranked 50 compounds resulted from non-covalent docking into the cathepsin B enzyme. For covalent docking, the nucleophilic center in the enzyme and the electrophilic sites of the ligands need to be defined. For Cathepsin B, Cys29 was assigned to be the nucleophilic center as evident from previous studies (*2, 6, 8*). CovalentDock

automatically identifies the electrophilic sites of the mined ligands - provided that they should be in the correct mol2 format. The binding site center corresponding to the active site Cys29 was X=28.8, Y=37.2 and Z=37.2. The docked 50 ligands were subsequently ranked according to the covalent binding energies.

Validation of docking approach

To validate the covalent docking approach used in this study, the same docking procedure has been applied to a set of experimentally determined compounds with known anti-cathepsin B activity obtained from the PubChem bioassay database (http://pubchem.ncbi.nlm.nih.gov/pcassay) (PubChem Bioassay ID: 820).

2.4. Molecular dynamics simulations

Molecular dynamics simulations were performed for the top-ranked covalently docked ligand-enzyme complex using the Amber12 software package (*18*). The full scenario of molecular dynamics simulations is described in our previous reports (*19-22*), however, a brief description is provided herein.

Owing to the lack of parameters needed for the ligand in the Cornell et al. force field (*23*), the missing parameters were created. Optimization of the ligands are first performed at the HF/6-31G* level with the Gaussian 03 package (*24*). The restrained electrostatic potential (RESP) procedure (*25*), is used to calculate the partial atomic charges. GAFF (*26*) force field parameters and RESP partial charges are assigned using the ANTECHAMBER module in the Amber12 package. Hydrogen atoms of the proteins were added using the Leap module in

Amber12 (*18*). The standard AMBER force field for bioorganic systems (ff03) (*27*) was used to define the enzyme parameters. Counter ions were added to neutralize the charge enzyme. The system was enveloped in a box of equilibrated TIP3P (*28*) water molecules with 8 Å distance around the enzyme. Cubic periodic boundary conditions were imposed and the long-range electrostatic interactions were treated with the particle-mesh Ewald method (*29*) implemented in Amber12 with a non-bonding cut-off distance of 10 Å. Initial energy minimization, with a restraint potential of 2 kcal/molÅ^2 applied to the solute, was carried out using the steepest descent method in Amber12 for 1000 iterations followed by conjugate gradient protocol for 2000 steps. The entire system was then freely minimized for 1000 iterations. Harmonic restraints with force constants 5 kcal/mol Å^2 were applied to all solute atoms during the heating phase. A canonical ensemble (NVT) MD was carried out for 50 ps, during which the system was gradually annealed from 0 to 300 K using a Langevin thermostat with a coupling coefficient of 1/ps. Subsequently, the system was equilibrated at 300 K with a 2 fs time step for 100 ps while maintaining the force constants on the restrained solute. The SHAKE algorithm (*30*) was employed on all atoms covalently bonded to a hydrogen atom during equilibration and production runs. With no restraints imposed, a production run was performed for 2 ns in an isothermal isobaric (NPT) ensemble using a Berendsen barostat (*31*) with a target pressure of 1 bar and a pressure coupling constant of 2 ps. The coordinate file was saved every 1 ps and the trajectory was analyzed every 1 ps using the Ptraj module implemented in Amber12.

3. Results and Discussion

3.1. Hybrid Non-covalent/covalent docking
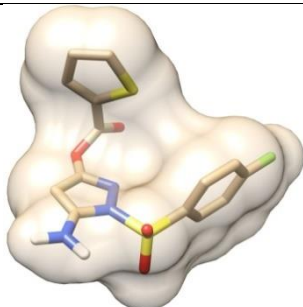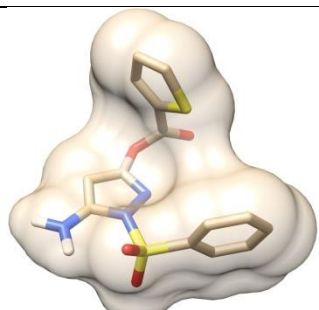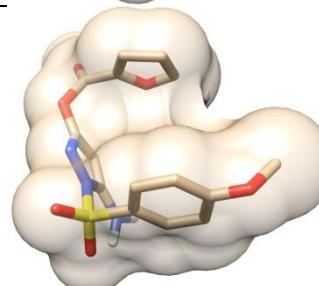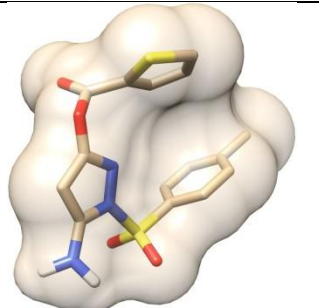
Since this work was primarily aimed to investigate covalent enzyme inhibition, covalent docking would be the most appropriate tool to provide better description of the binding theme of this class of inhibitors. However, due to the fact that covalent docking is time consuming, virtual screening of large compound library using covalent docking approach, would be unfeasible. To this end, we opted to run initial non-covalent docking to rank the 1186 compounds according to their binding affinities towards the target protein, and then only the best 50 docked compounds will be selected for a subsequent covalent docking.

Validation of docking calculations
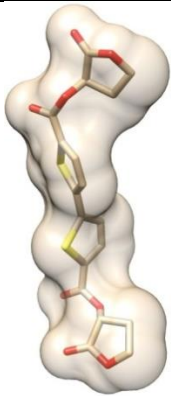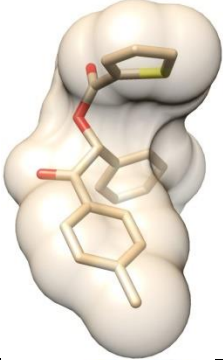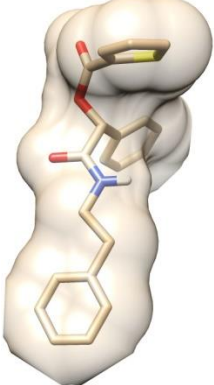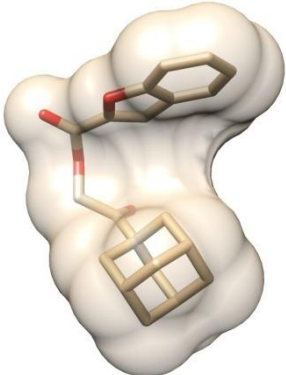
To assess the reliability and validity of the covalent docking approach, the same docking protocol was performed on a set of compounds with available experimental data (IC50). The PubChem codes of these compounds are CID3241895, CID3243128, CID646525 and CID653862. As evident from Table 1, the docking results, to a great extent, are consistent with the experimental data - this proves that the covalent docking method applied in this study is reasonably reliable.
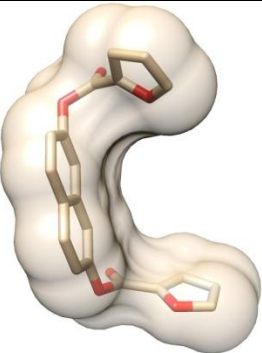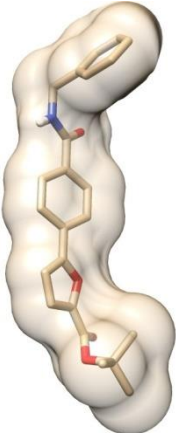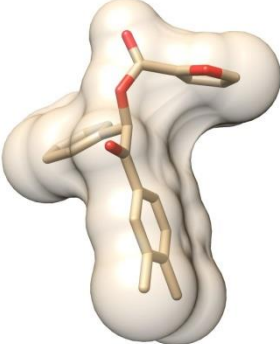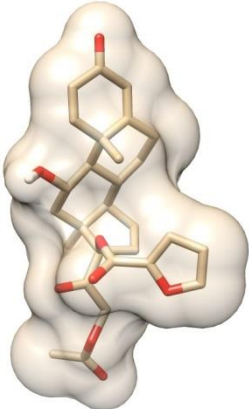
The best 10 docked structures resulted from covalent docking are shown in Table 1. Also, as evident from Table 1, these compounds do not violate Lipinski's rule of five, in turn, are likely to be orally bioavailable.
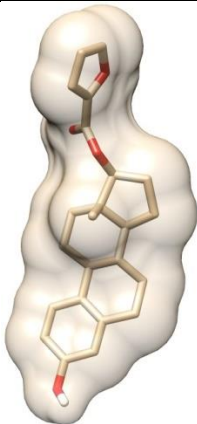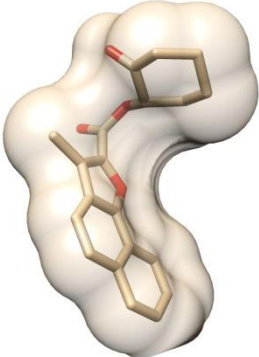
Table 1: The ID codes, 3D structures, covalent dock binding energies and Lipinski's rule of five for the experimental compounds and top-10 covalently docked compounds.

Compound Code	3D Structure	*ΔG (kcal/ mol)	LogP	M.W (g/mol)	Hydrogen Bonds	
					Donor	Acceptor
Experimental Compounds (PubChem codes)						
CID3241895 IC$_{50}$: 0.44		-11.6	3.50	367.375	1	8
CID3243128 IC$_{50}$: 0.25		-11.3	3.40	349.385	1	7
CID646525 IC$_{50}$: 1.99		-11.1	3.80	363.411	1	7
CID653862 IC$_{50}$: 0.92		-11.0	2.80	363.345	1	8

Compounds Library (ZINC codes)

ZINC03378824		-12.3	1.11	422.436	0	8
ZINC00477774		-11.9	4.85	336.412	0	3
ZINC04901668		-11.7	4.16	365.454	1	4
ZINC04332083		-11.5	4.44	338.403	0	4

ZINC00365205		-11.0	4.67	348.310	0	6
ZINC60246405		-10.9	4.72	377.440	1	5
ZINC00462034		-10.6	4.58	334.371	0	4
ZINC29347283		-9.9	3.65	496.556	1	8

ZINC15204533		-9.0	4.78	366.457	1	0
ZINC07014955		-8.1	4.44	322.360	0	4

*CovalentDock estimated binding energy

Interestingly, compounds ZINC03378824, ZINC04332083 and ZINC00477774 showed better binding affinities towards cathepsin B when compared to prototypes (CID3241895, CID3243128, CID646525 and CID653862) (see Table 1). We also noted that the two top-ranked compounds, ZINC03378824 and ZINC00477774 are thiophene-based inhibitors rather than furan-based – this is in a great accordance with experimental data (Table 1). The retrieved compounds library did not show any pyrrole-based inhibitors. Furthermore, the docked binding energies for the experimentally determined compounds (obtained from PubChem bioassay database) correlate well with the IC$_{50}$ (Table 1). All these findings verify that the adopted method is, to a great extent, reproducible and reliable.

3.2. Molecular dynamic and post-dynamic analysis

From our experience with molecular docking, in many occasions, we observed that even best docked structures are not stable enough in the enzyme active site after a short run of MD simulations. Therefore, we believe that, it is always important to run a reasonably long MD simulation to ensure that the docked complexes are stable and not docking artifacts.

For the best-docked structure, ZINC03378824-enzyme complex, we ran 2 ns MD simulations. As evident from Figure 3, all-atom RSMD and the total potential energy fluctuations clearly verify that the system is quite stable over the MD simulations and indicate that the interactions with the ligand site were strong enough to maintain the ligand bonding.

The average complex structure (Figure 4) was extracted over the 2 ns MD using ptraj module in Amber software (*32*) and then subjected to further per-residue interaction energy decomposition and binding forces analysis.
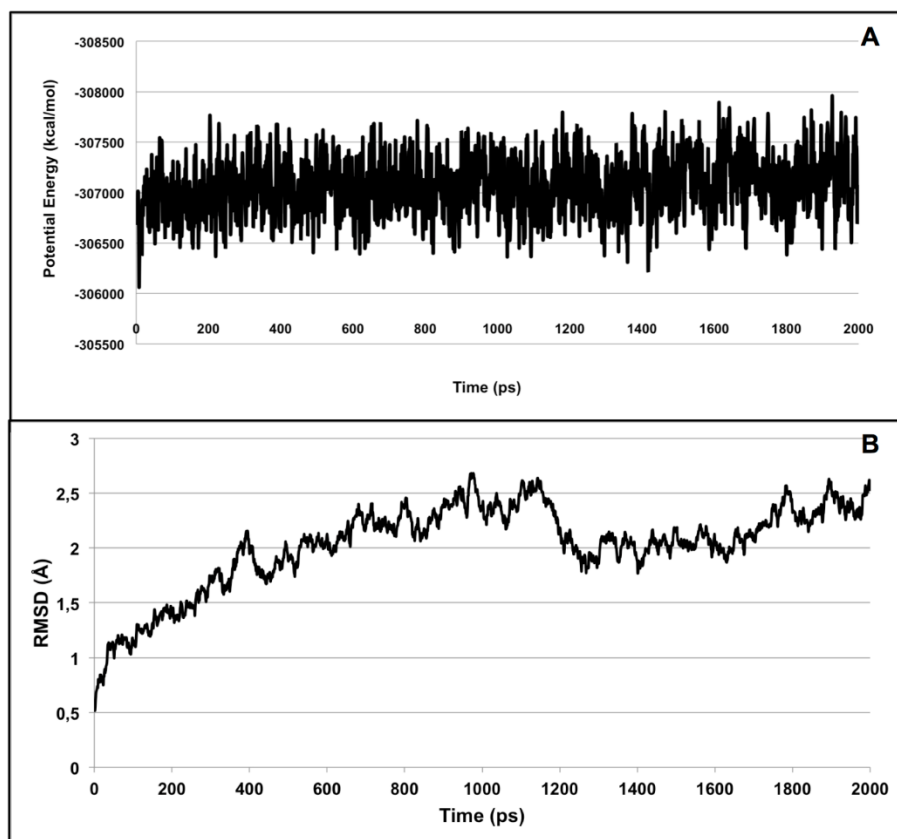
Figure 3: Potential energy and RMSD, A and B, respectively, for the MD trajectory for the ZINC03378824-enzyme complex.
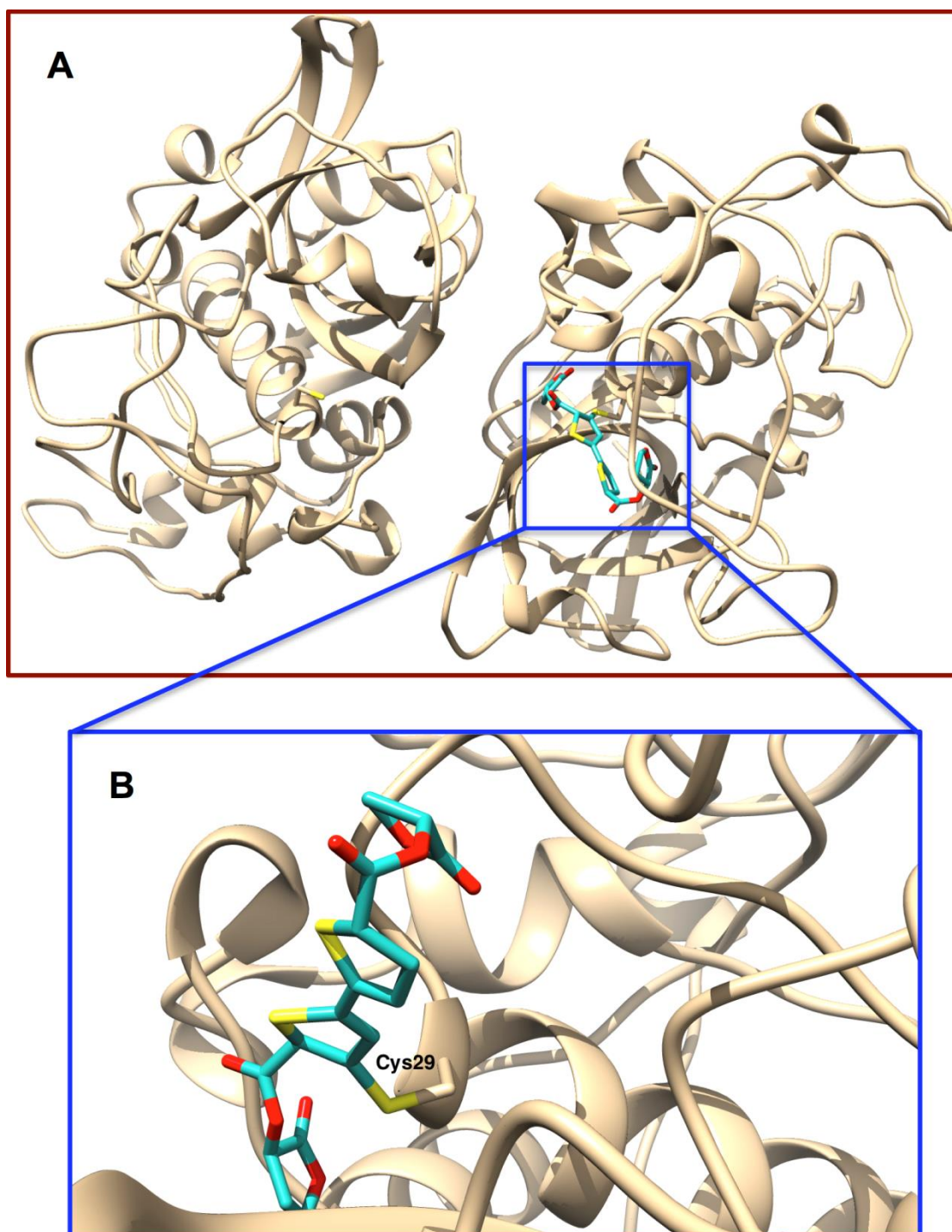
Figure 4: 2ns MD average structure of the top-ranked compound, ZINC03378824, covalently bonded to active site of cathepsin B: (A) 3D representation showing the ligand at the enzyme dimer interface and (B) a closer view showing the ligand covalently bonded to the active site Cys29.

To further explore the crucial active site amino acid residues contributed to the binding of the ligand, we estimated the per-residue interaction energy decomposition using the MolDock scoring function (*33*). This function allowed us to investigate the effect of each amino acid that contributed in the interactions observed between the enzyme and the hit compounds.
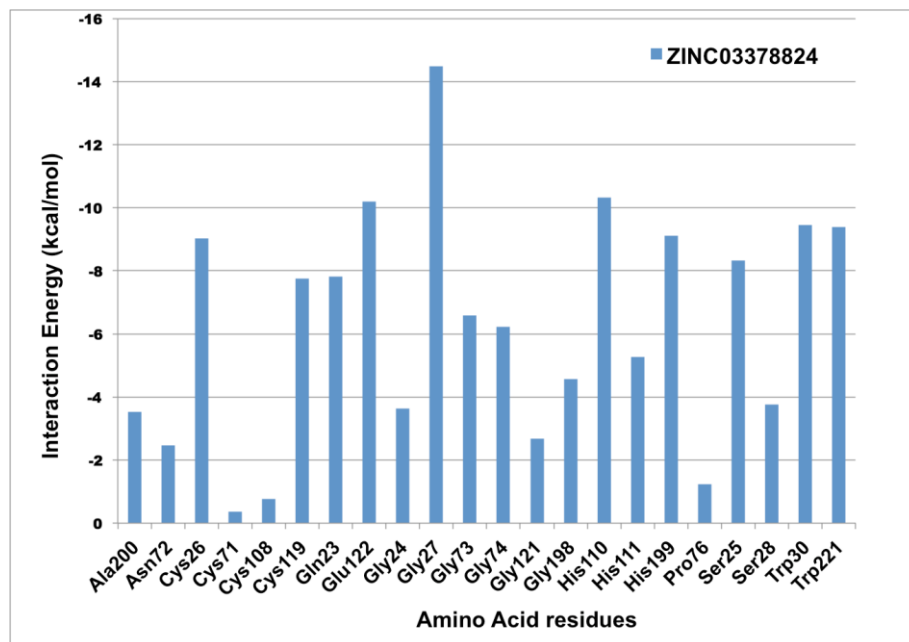
Figure 5: Per-residue energy decomposition for the top-ranked compound (ZINC03378824) using the MolDock scoring function.

It was observed from Figure 5 that amino acids Cys26, Cys119, Gln23, Glu122, Gly27, His110, His199, Ser25, Trp30 and Trp221 are interacting the most with the ligand. To provide better insight into the nature of the binding forces between the above-mentioned amino acids and the ligand, we analyzed the binding forces using MOE software (www.chemcomp.com) (Figure 6).
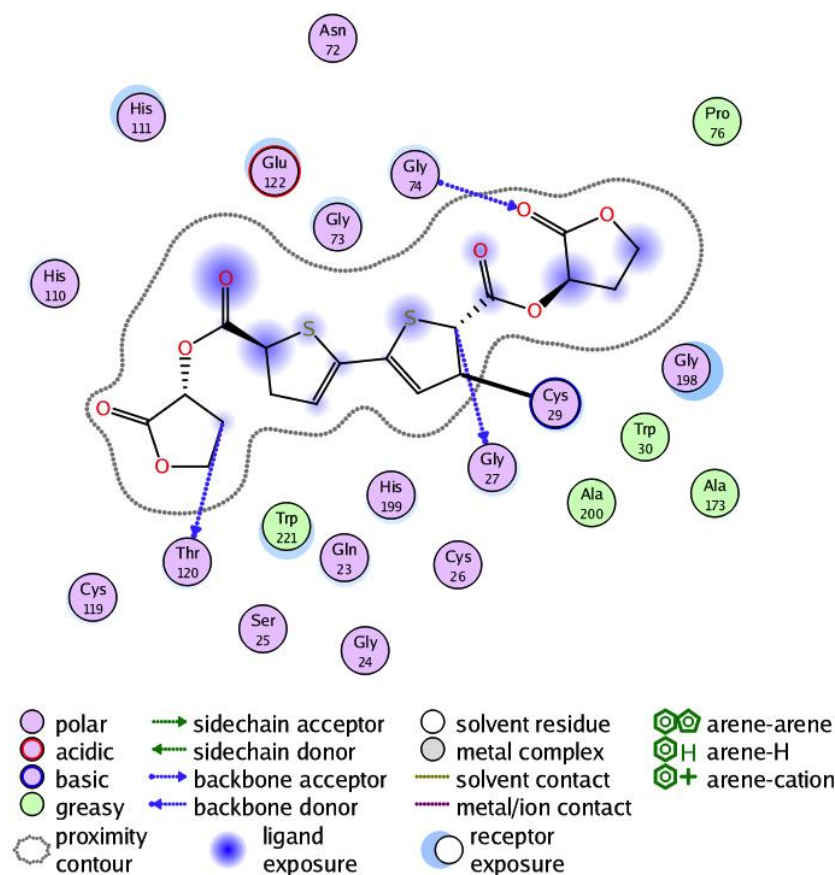
Figure 6: Different binding forces between the enzyme active site residues and the best hit, ZINC03378824.

As evident from Figure 6, polar interactions are the most prominent binding forces that contributed to the binding of the ligand. As shown in Figure 5 and 6, residues involved in such interactions are Cys26, Cys119, Gln23, Glu122, Gly27, His110, His199, Ser25, Trp30 and Trp221. Hydrophobic interactions between the ligand and Ala200, Ala173, Trp30, Trp221 and Pro76 were also observed. Information gained from per-residue interaction decomposition and binding forces analyses, Figure 5 and 6, respectively, should assist in the design of ligands that can form favorable interaction with the aforementioned active site residues.

In light of agreement of the computational results with the experimental findings, the unique covalent docking approach adopted in this work supported by MD simulation has proved successful in the prediction of the binding themes and affinities of cathepsin B covalent inhibitors.

The novelty of the four hit compounds, ZINC03378824, ZINC00477774, ZINC04901668 and ZINC04332083 was further confirmed by the SciFinder Scholar search as these compounds have not been experimentally reported for cathepsin B inhibitory activity earlier. Hence, we suggest that these four compounds are novel scaffolds as cathepsin B inhibitors and could be potential anti-cancer drugs.

Despite the fact that the integrated computational workflow presented in this work could serve as time-efficient powerful tool in the process of drug design and discovery of covalent inhibitors, further experimental testing of the virtual hits would be confirmatory to our findings.

4. Conclusion

In the present study, we report a unique workflow involving hybrid ligand/structure based virtual screening using covalent docking approach to search for novel classes of irreversible cathepsin B inhibitors based on Michael acceptors-like structures. Furthermore, the current work has provided insight into the active site residues involved in ligand binding through per-residue interaction energy decomposition and binding forces analysis. Results showed that polar interactions between the ligand and active site residues have contributed the most towards the ligand binding. Also, hydrophobic interactions played a role in drug binding.

However, further experimental investigations are still required to support these computational findings, we believe that the various *in-silico* tools demonstrated in this report should serve as an initial step that could assist medicinal chemists to design more potential irreversible inhibitors.

Acknowledgment

Supplementary material

The PDB coordinates for all docked structures are provided as supplementary data

References

1.   Shah Unnati, S. R., Acharya Sanjeev, Acharya Niyati. (2013) Novel anticancer agents from plant sources, *Chinese Journal of Natural Medicines 11*, 0016−0023.

2.   Zhou, Z. G., Wang, Y. L., and Bryant, S. H. (2009) Computational Analysis of the Cathepsin B Inhibitors Activities Through LR-MMPBSA Binding Affinity Calculation Based on Docked Complex, *J. Comput. Chem. 30*, 2165-2175.

3.   Yu Lin, W. J. W. (1996) Molecular modeling of substrate-enzyme reactions for the cysteine protease papain, *J. Mol. Graphics 14*, 62-72.

4.   Zhou, Z., Wang, Y., and Bryant, S. H. (2011) Multi-conformation 3D QSAR study of benzenesulfonyl-pyrazol-ester compounds and their analogs as cathepsin B inhibitors, *J. Mol. Graph. Model. 30*, 135-147.

5.   Tomoo, K. (2010) Development of Cathepsin Inhibitors and Structure-Based Design of Cathepsin B-Specific Inhibitor, *Current Topics in Medicinal Chemistry 10*, 696-707.

6.   Sosic, I., Mirkovic, B., Turk, S., Stefaneb, B., Kos, J., and Gobec, S. (2011) Discovery and kinetic evaluation of 6-substituted 4-benzylthio-1,3,5-triazin-2(1H)-ones as inhibitors of cathepsin B, *Eur. J. Med. Chem. 46*, 4648-4656.

7.   Zhou, Z., Wang, Y., and Bryant, S. H. (2010) QSAR models for predicting cathepsin B inhibition by small molecules--continuous and binary QSAR models to classify cathepsin B inhibition activities of small molecules, *J. Mol. Graph. Model. 28*, 714-727.

8.   Sosic, I., Mirkovic, B., Arenz, K., Stefane, B., Kos, J., and Gobec, S. (2013) Development of new cathepsin B inhibitors: combining bioisosteric replacements and structure-based design to explore the structure-activity relationships of nitroxoline derivatives, *J. Med. Chem. 56*, 521-533.

9.   Santos, M. M. M., and Moreira, R. (2007) Michael acceptors as cysteine protease inhibitors, *Mini-Rev. Med. Chem. 7*, 1040-1050.

10.   Eckert, H., and Bojorath, J. (2007) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches, *Drug Discov. Today 12*, 225-233.

11.   Heikamp, K., and Bajorath, J. (2013) The Future of Virtual Compound Screening, *Chem. Biol. Drug Des. 81*, 33-40.

12.   Walters, W. P., Stahl, M. T., and Murcko, M. A. (1998) Virtual screening - an overview, *Drug Discov. Today 3*, 160-178.

13.   Willett, P. (1998) Chemical Similarity Searching, *Journal of Chemical Information and Computional Science 38*, 983-996.

14.   Ouyang, X., Zhou, S., Su, C. T. T., Ge, Z., Li, R., and Kwoh, C. K. (2013) CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints, *J. Comput. Chem. 34*, 326-336.

15.   Forli, S. AutoDock | Raccoon: an automated tool for preparing AutoDock virtual screenings.

16.   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, *- 31*, - 461.

17.   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, *Journal of Computational Chemistry 25*, 1605-1612.

18.   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, *Journal of Computational Chemistry 26*, 1668-1688.

19.   Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E., Sayed, Y., Ibrahim, M. A., Naicker, P., and Soliman, M. E. (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.

20.   Karpoormath, R., Sayed, Y., Govender, P., Govender, T., Kruger, H. G., Soliman, M. E., and Maguire, G. E. (2012) Pentacycloundecane derived hydroxy acid peptides: a new class of irreversible non-scissile ether bridged type isoster as potential HIV-1 wild type C-SA protease inhibitors, *Bioorg. Chem. 40*, 19-29.

21.   Makatini, M. M., Petzold, K., Arvidsson, P. I., Honarparvar, B., Govender, T., Maguire, G. E., Parboosing, R., Sayed, Y., Soliman, M. E., and Kruger, H. G. (2012) Synthesis, screening and computational investigation of pentacycloundecane-peptoids as potent CSA-HIV PR inhibitors, *Eur. J. Med. Chem. 57*, 459-467.

22.   Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. A., Dirr, H. W., Soliman, M. E., and Sayed, Y. (2012) Structural insights into the South African HIV-1

subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance, *J. Biomol. Struct. Dyn.*

23. 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, *Journal of the American Chemical Society 117*, 5179-5197.

24. 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) *Gaussian Inc, Wallingford, CT*.

25. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins, *J. Comput. Chem. 16*, 1357-1377.

26. Wang J Fau - Wolf, R. M., Wolf Rm Fau - Caldwell, J. W., Caldwell Jw Fau - Kollman, P. A., Kollman Pa Fau - Case, D. A., and Case, D. A. Development and testing of a general amber force field.

27. 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, *Journal of Computational Chemistry 24*, 1999–2012.

28. 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.

29. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.

30. Ryckaert, J. P. C., G.; Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys 23*, 327–341.

31. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684–3690.

32. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.

33. Thomsen, R., and Christensen, M. H. (2006) MolDock: A new technique for high-accuracy molecular docking, *J. Med. Chem. 49*, 3315-3321.

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*]

[a]School 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 IC_{50} 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 modelling 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 5ns 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

5.1. 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 1.
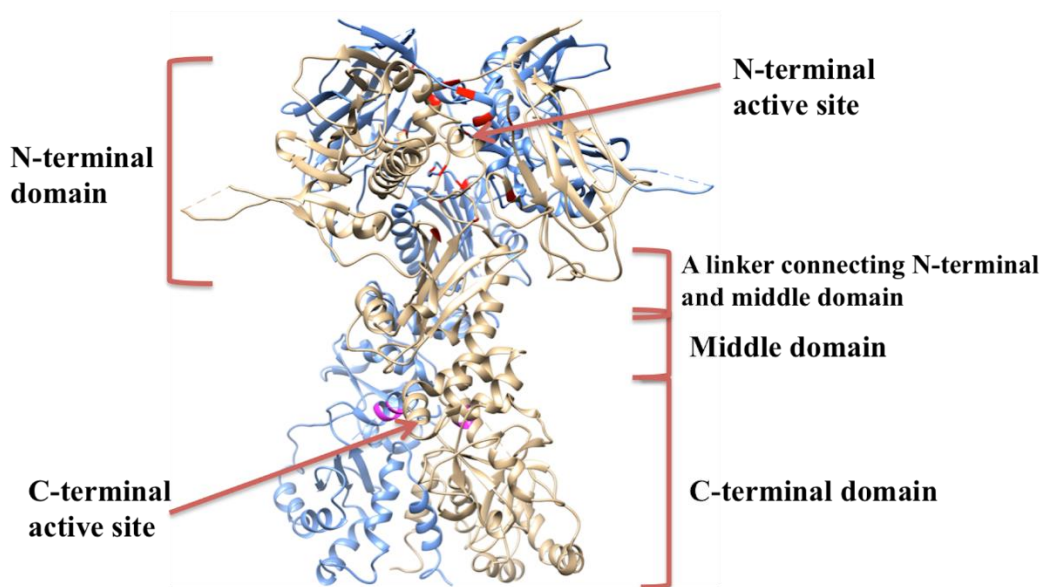
Figure 5.1: The crystal structure of Hsp90 Alpha (blue) and Beta chain (gold) (PDB code: 2CG9) showing its different domains (*15*).

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 Hsp90α 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 (NFV) mesylate, geftinib, 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 PIs 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 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 (5ns 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.
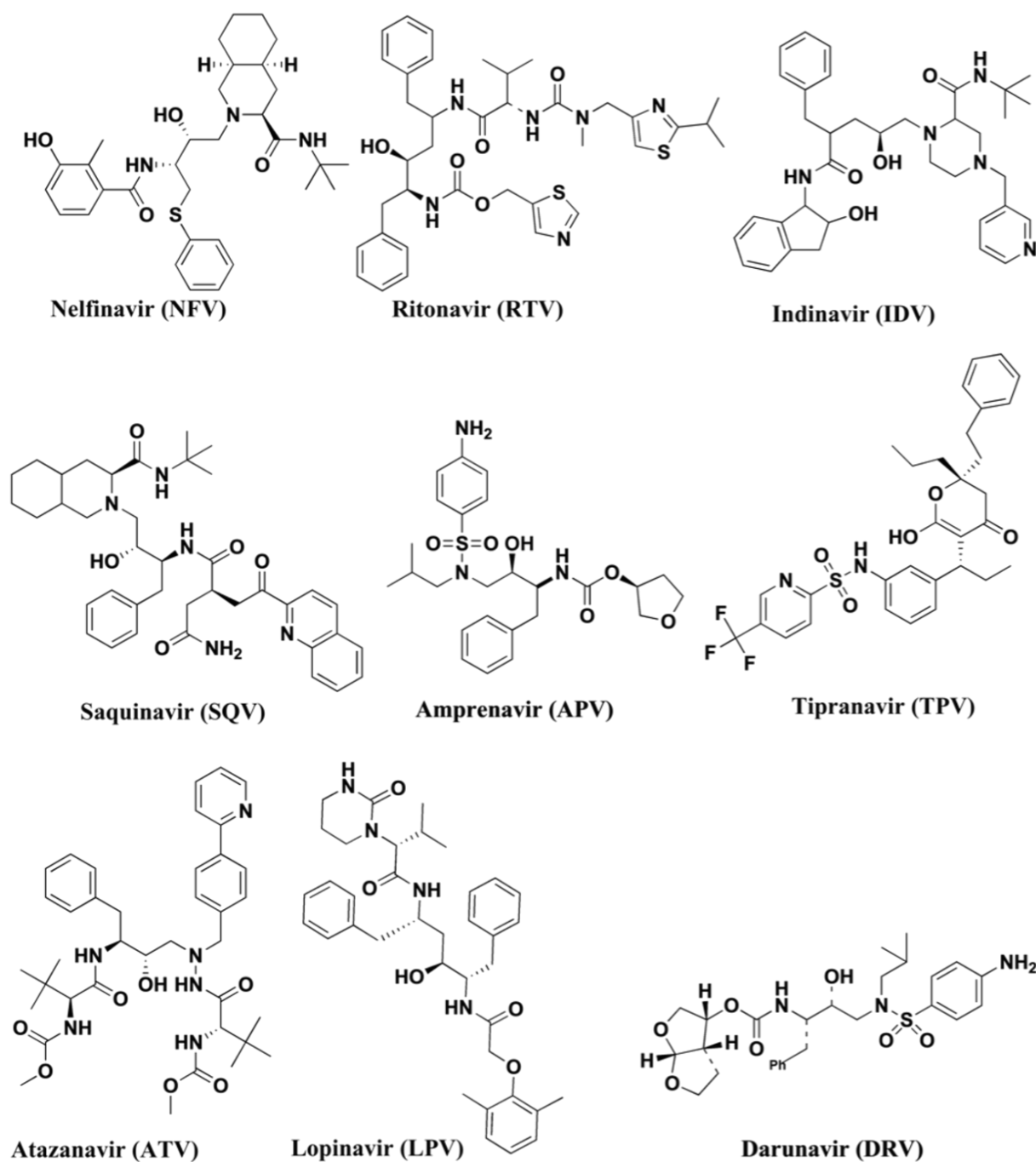
Figure 5.2: A schematic representation of 2D structures of the nine FDA-approved HIV-1 protease inhibitors.

5.2. Computational Methodology

5.2.1. Homology modeling of human Hsp90 protein structure

Due to the absence of a full human crystal structure of Hsp90 that comprises 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).

5.2.2. 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 Hsp90β homologue to identify the NTD and CTD active site pockets for further docking and molecular dynamic simulations.

5.2.2.1. 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.

5.2.2.2. 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.
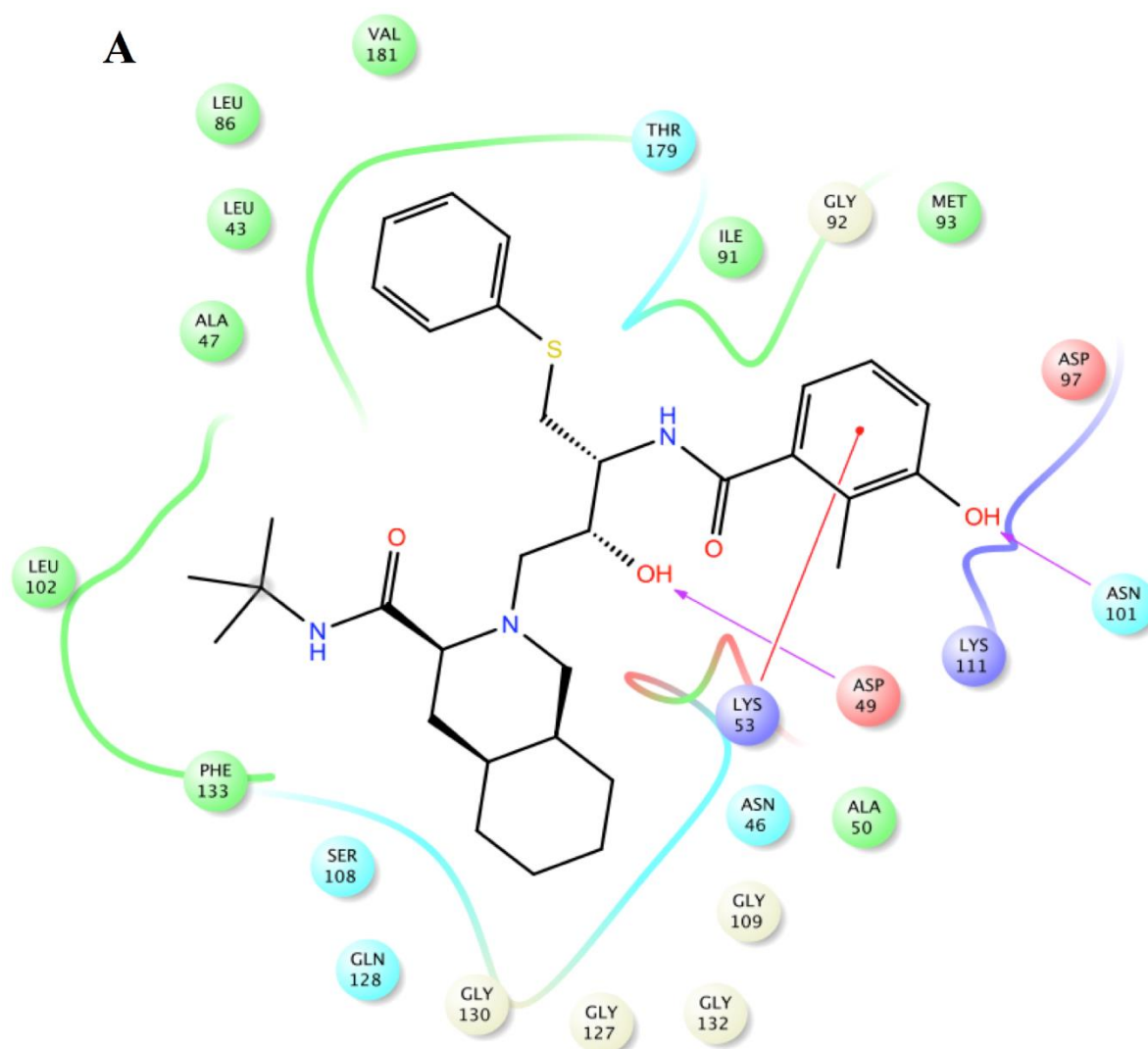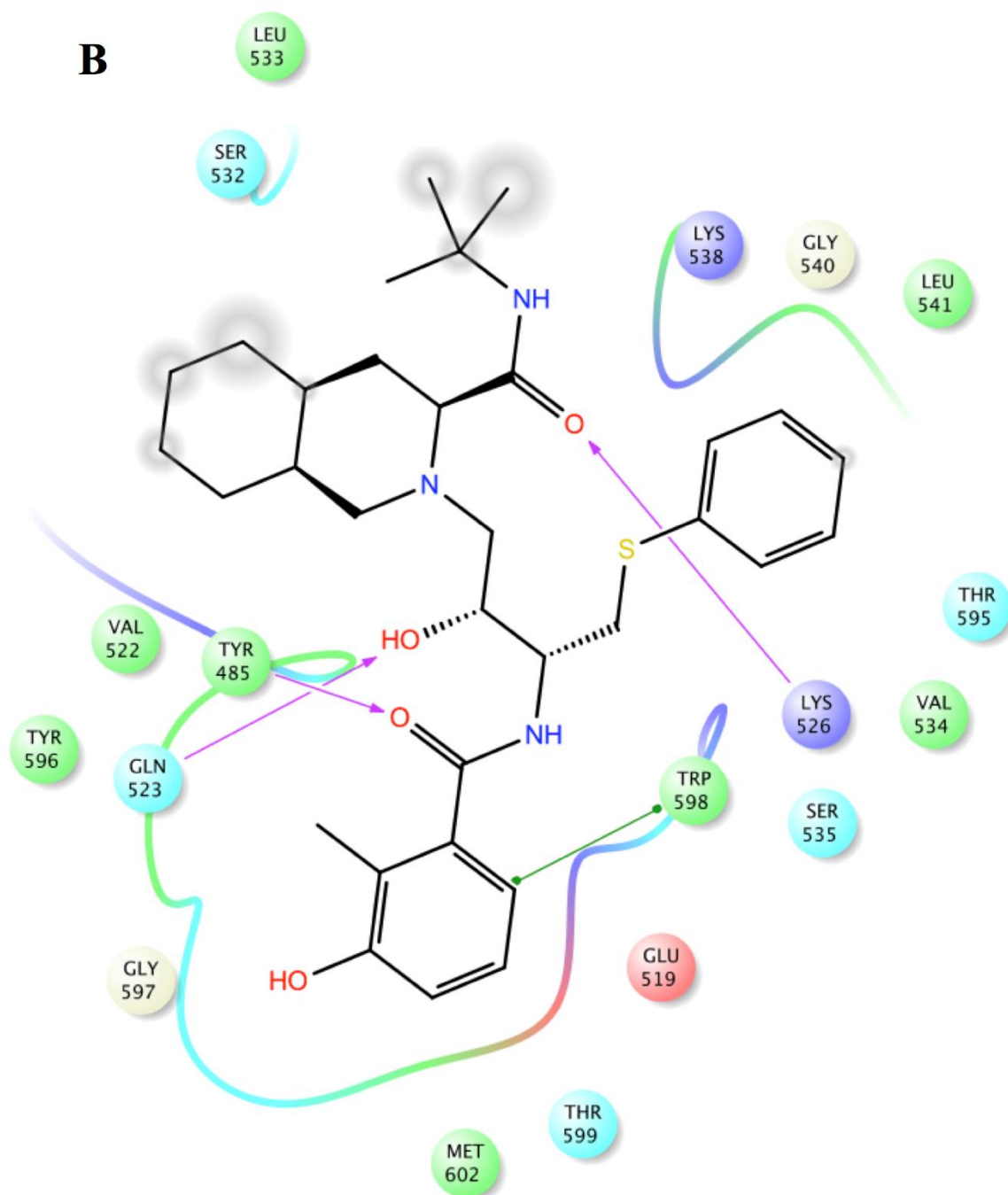
Figure 5.3: (A) Nelfinavir-human Hsp90 homologue interaction at the NTD (B) Nelfinavir-human Hsp90 homologue interaction at the CTD. Major contributions are from the residues that exhibit hydrophobic interactions (in green bubbles). Each illustration for other ligands at both terminals is provided in the supplementary material-S3.

5.2.3. 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.

5.2.4. 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/molÅ^2 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 Å^2 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 5ns 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.

5.2.5. 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:

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

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

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

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

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

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·Å^2) (*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*).

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

5.3. Results and Discussion

5.3.1. The Human Hsp90 homology model

In an effort to work with a human Hsp90 that is comprised of the middle domain (MD), C-terminal domain (CTD) and N-terminal domain (NTD), a homology model of the human Hsp90β 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 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 4b).

A   **B**
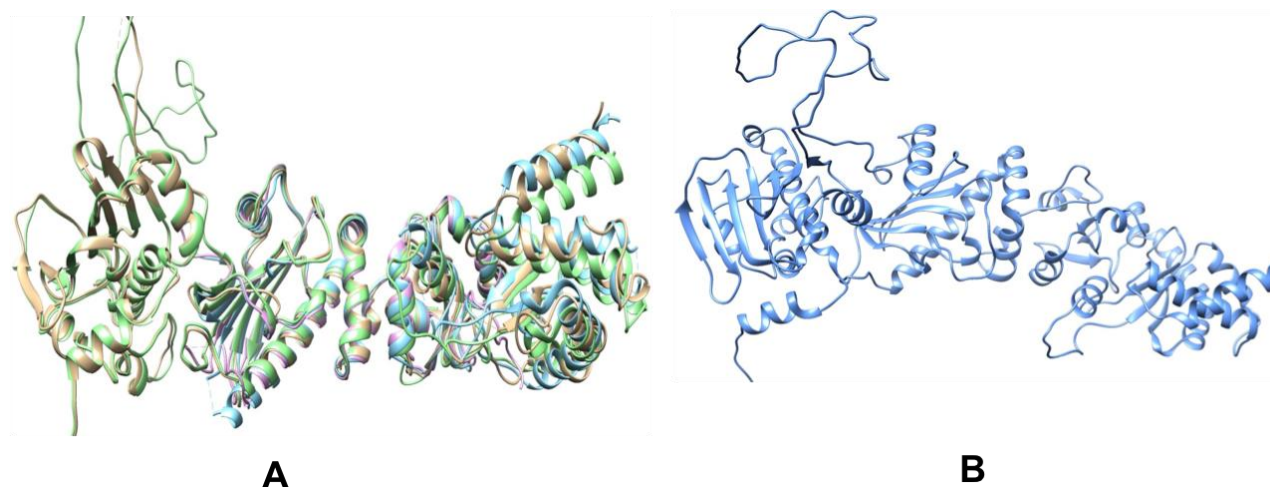
Figure 5.4: (A) Superimposed structures of the 2CG9 (gold), 3PRY (purple) and 3HGC (blue) and the generated human homologue sequence (green), (B) Generated homology model of the human Hsp90β.

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 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*).

NFV at the N-terminal Domain

NFV at the C-terminal Domain

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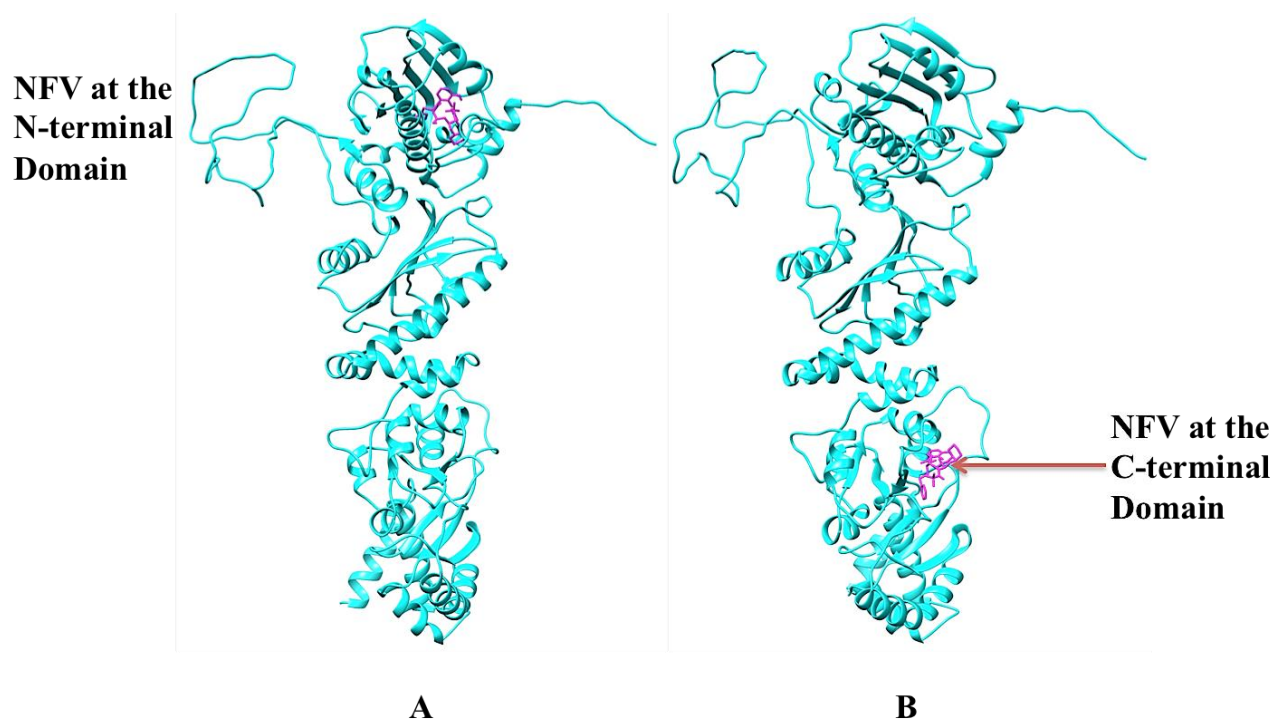
Figure 5.5: The human Hsp90 homologue docked with NFV at (a) N-terminal domain and (b) C-terminal domain.

5.3.2. Molecular Dynamics (MD) simulations

The structural dynamics of all HIV-1 PIs, which are bound to the human Hsp90 homologue were analysed by performing 5ns 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.

5.3.2.1.   Post-dynamic analyses:

Stability of Molecular Dynamics simulation

As shown in figure 6, the system was well equilibrated where the RMSD value of all nine complexes at the NTD did not exceed 2Å and stabilised after 1000 ps. On the other hand, the CTD shows that all the nine complexes did not exceed 2Å and stabilised after 1800 ps. 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 $\leq$ 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.
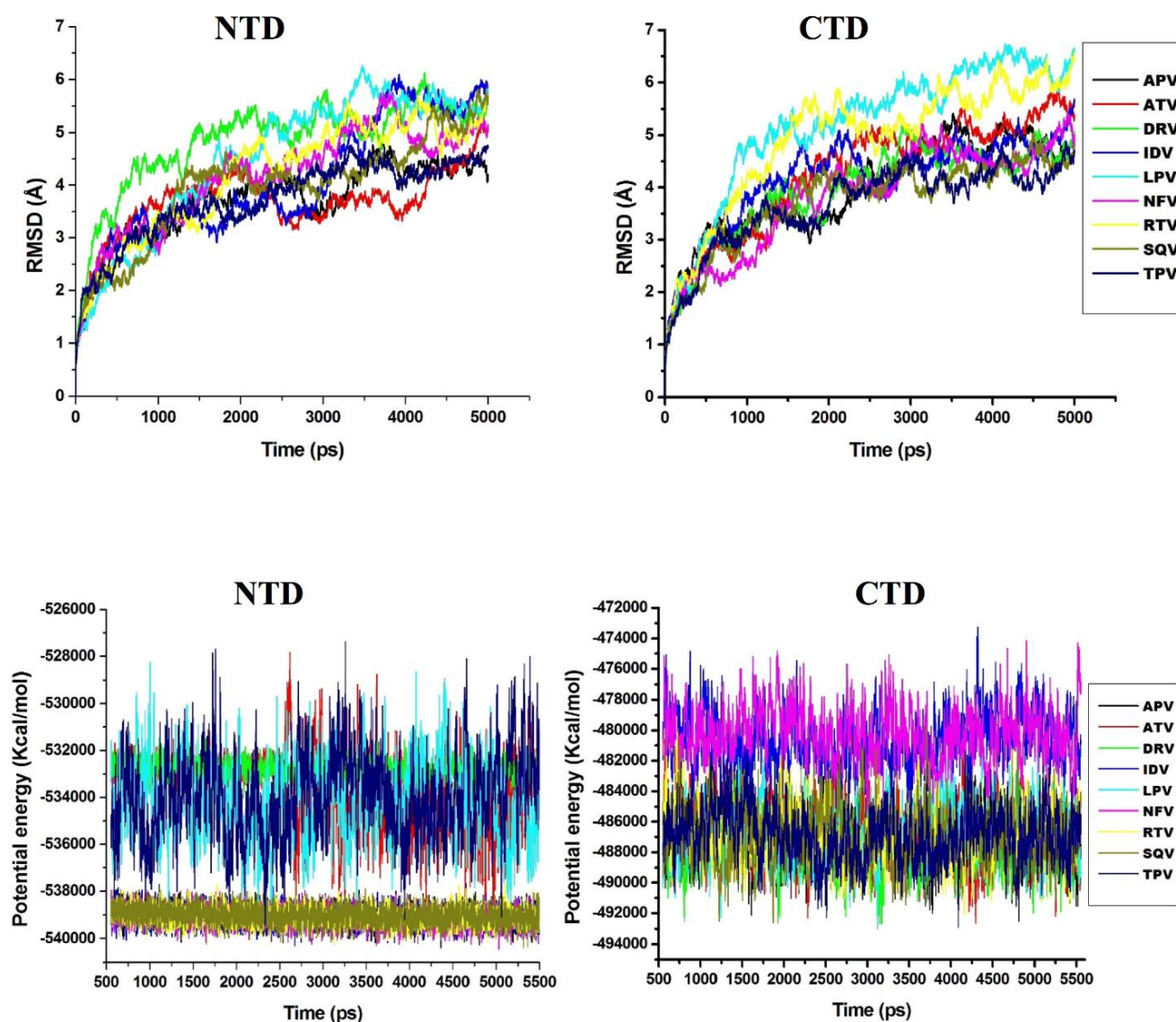
Figure 5.6: Comparative RMSD and potential energy plot of all ligands binding at the N-terminal domain and C-terminal domain respectively. (Individual plot for each ligand is provided in the supplementary material).

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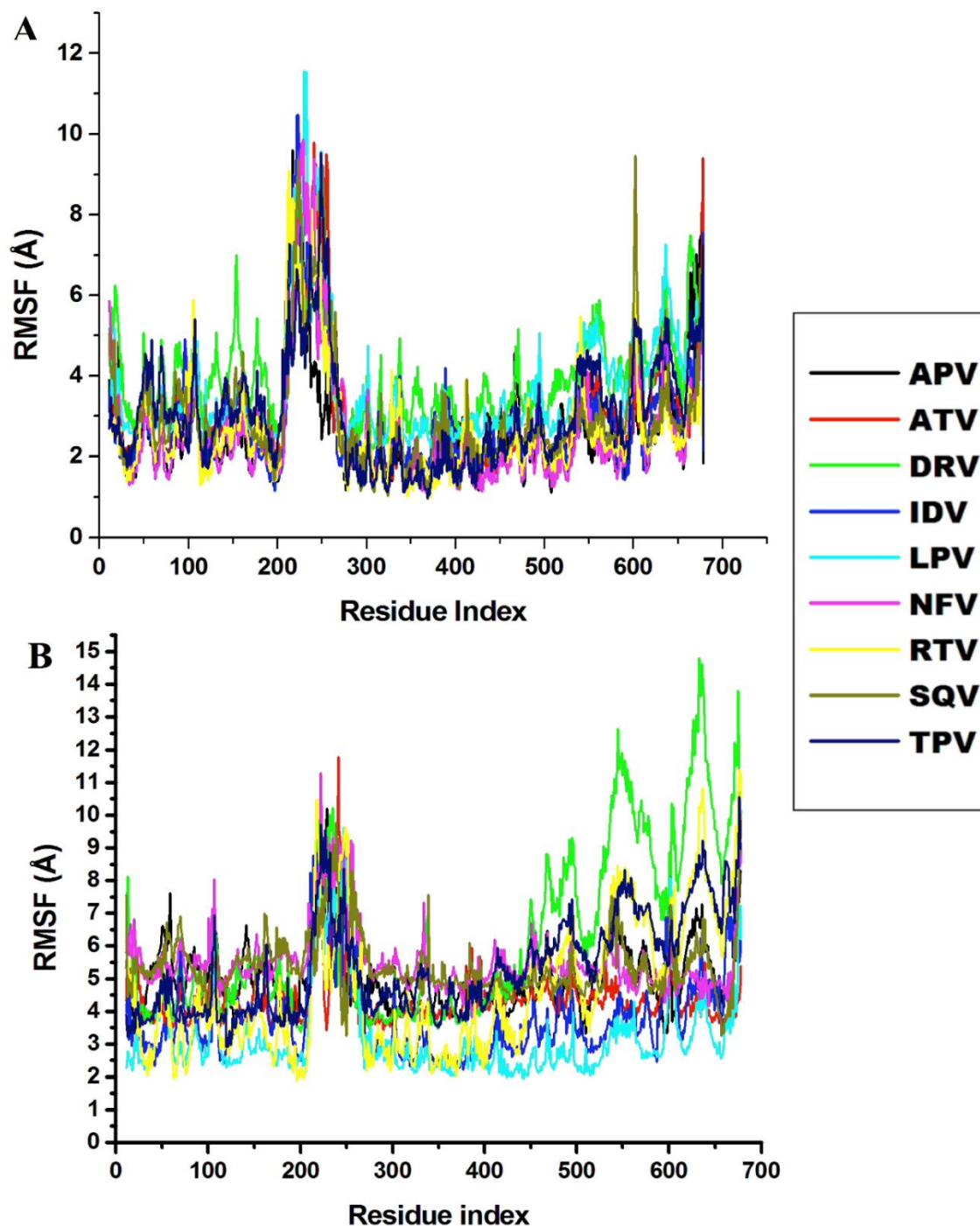
Figure 5.7: (A) Comparative RMSF plot of all ligands binding at the N-terminal domain and (B) Comparative RMSF plot of all ligands binding at the C-terminal domain. (Individual plot for each ligand is provided under the supplementary material).

5.3.2.2.   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 5ns 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 **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 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	-57.7±0.1	-7.72±0.1	100.84±1.4	-25.52±1.3	108.56±1.4	-83.03±0.5
	-103.0±0.9	*-51.9±0.9*	*-7.3±0.1*	*128.9±1.2*	*-160.9±1.7*	*121.7±1.1*	*-39.3±0.6*
RTV	-14.9±0.3	-80.1±0.3	-10.7±0.02	40.8±0.3	-94.9±0.5	30.1±0.3	-64.8±0.3
	-17.1±0.4	*-76.2±0.4*	*-9.2±0.04*	*38.6±0.4*	*-93.3±0.9*	*29.4±0.4*	*-63.9±0.4*
SQV	-273.3±1.0	-89.5±0.3	-10.7±0.01	*308.4±0.9*	-362.8±1.1	297.7±0.9	-64.8±0.4
	-97.7±1.9	*-38.4±1.0*	*-6.2±0.1*	*119.3±2.1*	*-136.1±2.8*	*113.1±1.9*	*-22.9±0.9*
IDV	-652.2±1.6	-81.3±0.4	-10.1±0.02	*676.2±1.5*	-733.4±1.6	666.1±1.5	-67.3±0.7
	-266.0±1.6	*-41.8±0.2*	*-5.4±0.05*	*303.6±1.6*	*-307.8±1.7*	*298.1±1.6*	*-9.7±0.4*
APV	-4.9±0.6	-67.5±1.7	-7.5±0.2	25.2±0.8	-72.4±1.9	17.7±0.7	-54.7±1.4
	-7.9±0.3	*-48.9±0.5*	*-6.4±0.04*	*22.2±0.3*	*-56.9±0.6*	*15.8±0.3*	*-41.1±0.5*
LPV	-8.0±0.7	-83.3±0.4	-9.7±0.04	30.7±0.6	-91.4±0.8	21.1±0.6	-70.3±0.5
	-12.2±0.3	*-56.5±0.4*	*-7.2±0.05*	*32.3±0.3*	*-68.8±0.5*	*25.1±0.3*	*-43.7±0.4*
ATV	-7.4±0.4	-89.8±0.4	-11.1±0.03	36.8±0.4	-97.2±0.6	22.8±0.4	-74.4±0.4
	-13.9±0.3	*-54.8±0.4*	*-7.0±0.04*	*36.3±0.3*	*-68.7±0.5*	*29.3±0.3*	*-39.4±0.4*
DRV	-16.9±0.5	-75.1±0.3	-9.2±0.03	36.8±0.5	-92.1±0.7	27.6±0.5	-64.5±0.5
	-12.9±0.3	*-47.1±0.3*	*-6.2±0.4*	*29.0±0.3*	*-60.0±0.4*	*22.9±0.3*	*-37.1±0.2*
TPV	-11.0±0.3	-77.6±0.3	-9.7±0.03	36.4±0.3	-88.6±0.41	26.7±0.3	-61.9±0.3
	-16.1±0.3	*-51.5±0.32*	*-6.7±0.03*	*35.5±0.3*	*-67.7±0.5*	*28.8±0.3*	*-38.9±0.3*

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

[b] $\Delta E_{gas} = \Delta E_{int} + \Delta E_{vdW} + \Delta G_{ele}$

[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 (IC$_{50}$ 10μM), 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 1, it is evident that nelfinavir binds readily to the human Hsp90 homologue NTD (ΔG_{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). 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 are required to further verify the binding mode of the human Hsp90 CTD to HIV-1 PIs.

5.3.2.3.   Per-residue interaction energy decomposition analysis

The binding free energy calculation was further performed to determine contributions from each human Hsp90β 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 8. These trends were maintained in all complexes at both terminals.
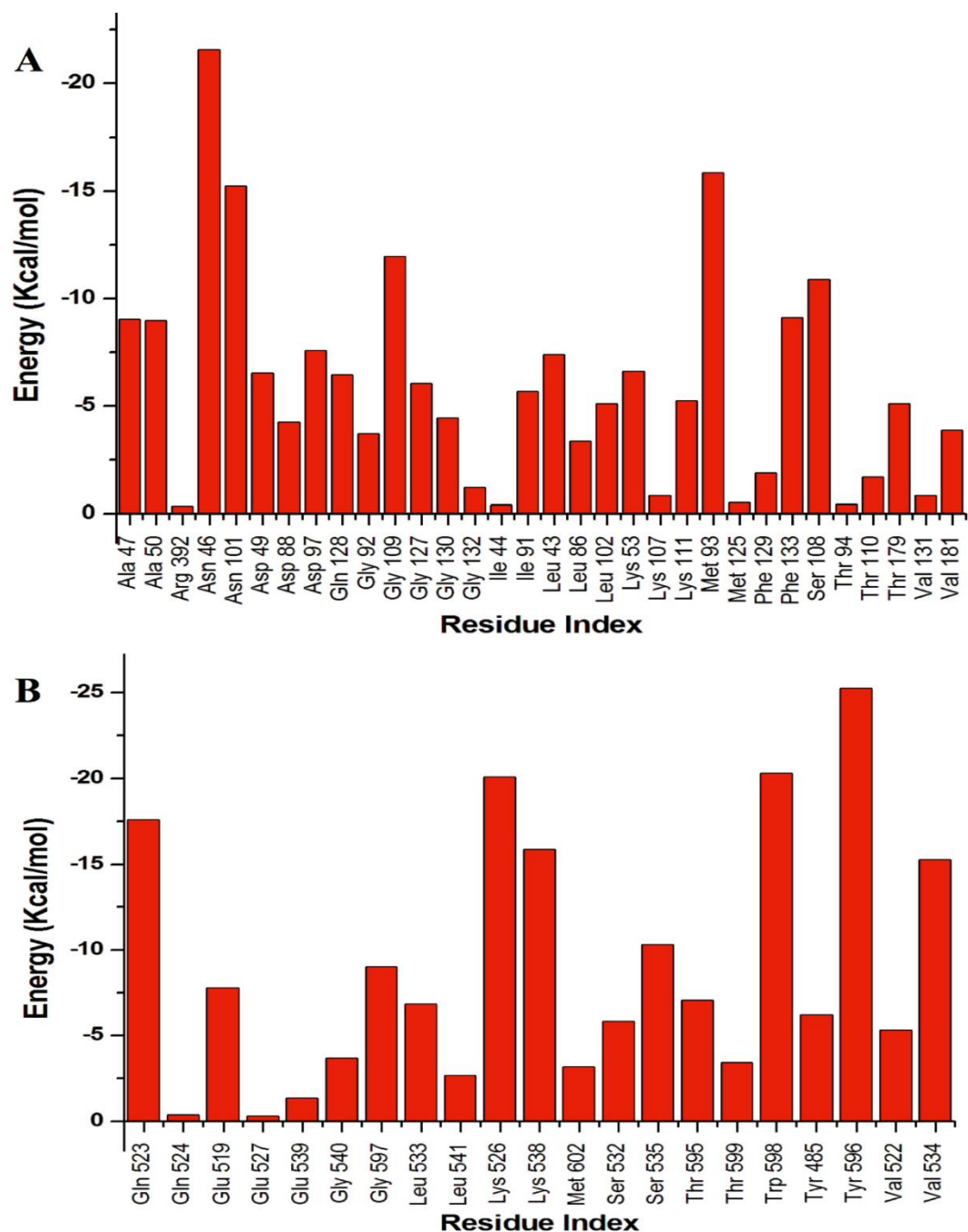
Figure 5.8: (A) Residues that contributed to the human Hsp90-nelfinavir binding at the NTD and (B) Residues contributing to the human Hsp90-nelfinavir binding at the CTD. (Illustrations for other ligands are provided in the supplementary material-S4).

5.4. Conclusion

In this study, the potential mechanism of binding of nine PIs to the human Hsp90β 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; MolProbity result including all validation data are provided in the supplementary material.

Acknowledgements

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, *Chinese Journal of Natural Medicines 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, *International Journal of 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 Detection and Prevention 30*, 118-128.

8. Stern, D. F. (2000) Tyrosine kinase signalling in breast cancer - ErbB family receptor tyrosine kinases, *Breast Cancer Research 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, *Science 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, *Journal of Clinical Oncology 20*, 719-726.

11. Gajria, D., and Chandarlapaty, S. (2011) HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies, *Expert Review of Anticancer Therapy 11*, 263-275.

12. Kurebayashi, J. (2001) Biological and clinical significance of HER2 overexpression in breast cancer, *Breast cancer (Tokyo, Japan) 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. Perterson, 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_1222 4_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, *Molecular 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, *Nature Reviews: Drug Discovery 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, *Clinical Cancer Research 13*, 5183-5194.

19. Osborne, C. K. (1998) Steroid hormone receptors in breast cancer management, *Breast Cancer Research and Treatment 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, *Journal of the National Cancer Institute 104*, 1576-1590.

21. Srirangam, A., Mitra, R., Wang, M., Gorski, J. C., Badve, S., Baldridge, 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, *Clinical Cancer Research 12*, 1883-1896.

22.   Bernstein, W. B., and Dennis, P. A. (2008) Repositioning HIV protease inhibitors as cancer therapeutics, *Current Opinion in Hiv and 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, *Current Protocol in Protein Science*.

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, *Journal of Computational Chemistry 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, *Journal of Molecular Biology 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 Research 38*, W563-W568.

29.   Schrodinger. http://www.schrodinger.com/, 20 Oct. 2013.

30.   Swiss PDB. 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, *Current Topics in Medicinal Chemistry 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, *Chemistry & Biology 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, *Journal of Computational Chemistry 25*, 1605-1612.

35. Dym, O., Xenarios, I., Ke, H. M., and Colicelli, J. (2002) Molecular docking of competitive phosphodiesterase inhibitors, *Molecular Pharmacology 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, *Journal of Computational Chemistry 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, *Journal of Computational Chemistry 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), *Journal of the American Chemical Society 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, *Journal of Computational Chemistry 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, *Journal of Chemical Physics 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, *Journal of Computational Chemistry 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, *Journal of the American Chemical Society 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, *Journal of Computational Chemistry 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, *Journal of Computational Chemistry 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, *Journal of Chemical Physics 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, *Journal of Computational Physics 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, *Accounts of Chemical Research 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, *Perspectives in Drug Discovery and Design 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, *Journal of Physical Chemistry B 104*, 3712-3720.

54.   Sitkoff, D., Sharp, K. A., and Honig, B. (1994) Accurate calculation of hydration free-energies using macroscopic solvent models, *Journal of Physical Chemistry 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, *Computer Physics Communications 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, *Current Computer-Aided Drug Design 7*, 146-157.

57.   Morris, G. M., and Lim-Wilby, M. (2008) Molecular docking, *Methods in Molecular Biology 443*, 365-382.

58.   Durrant, J., and McCammon, J. A. (2011) Molecular dynamics simulations and drug discovery, *BMC Biology 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, *Chemical Biology and Drug design 81*, 208-218.

61.   Deng, Y., and Roux, B. (2009) Computations of Standard Binding Free Energies with Molecular Dynamics Simulations, *The Journal of Physical Chemistry 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, *Biophysical Journal 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, *Journal of Chemical Information and Modeling 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, *Organic & Biomolecular Chemistry 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, *Journal of Molecular Structure-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), *Journal of Medicinal Chemistry 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, *Journal of Molecular Biology 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, *Journal of Medicinal Chemistry 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, *Journal of the American Chemical Society 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, *Journal of Medicinal Chemistry 50*, 1768-1777.

72.	Perez, M. (2005) Gibbs-Thomson effects in phase transformations, *Scripta Materialia 52*, 709-712.

CHAPTER 6

This chapter presents general conclusions, recommendations as well as suggestions for future research.

6.1. General Conclusion

The main aim of this study was to investigate effective therapeutic drugs, using Cathepsin B and Hsp90 as our target enzymes that are involved in a variety of cancer pathological processes. The Michael acceptor-type compounds that show better covalent binding with Cathepsin B were investigated and drug binding modes and stability of the FDA-approved HIV-1 protease inhibitors to the target enzyme (Hsp90) binding sites so as to inhibit cancer were explored. The computational chemistry methods applied in this work provided a cost effective and an efficient tool for drug discovery and development. This research study has accomplished two practical aims of the study that resulted in the following conclusions:

1. A unique workflow that involved hybrid ligand/structure based virtual screening using covalent docking approach to search for novel irreversible cathepsin B inhibitors based on Michael acceptor-like structures was exploited. We discovered ZINC03378824 to have the highest covalent binding affinity with Cathepsin B active site (Cys29) and other compounds with better binding affinity when compared to the experimental compounds including ZINC00477774, ZINC04901668 and ZINC04332083.

2. The MD simulations result of the best hit compound (ZINC03378824) showed that all-atom RSMD is below 1 Å, and the variation in the total potential energy is less than 1000 kcal/mol, indicating that the interactions with the enzyme site were strong enough to maintain the ligand bonding, thus the system was quite stable. The amino acid residues that contributed the most to the binding of the top hit compound include Cys26, Cys119, Gln23, Glu122, Gly27, His110, His199, Ser25, Trp30 and Trp221. The binding forces of these amino acid residues were analysed, and it was observed that polar interactions were the most prominent binding forces. Residues involved in such interaction were Cys26, Cys119, Gln23, Glu122, Gly27, His110, His199, Ser25, Trp30 and Trp221. Hydrophobic interactions were also observed between the ligand and Ala200, Ala173, Trp30, Trp221 and Pro76 of the target enzyme. An integrated computational workflow introduced in

101

this study will serve as time-efficient powerful tool in the process of drug design and discovery of covalent inhibitors of Cathepsin B.

3. The relevance of relative affinities between human HSP90 enzyme and HIV-1 protease inhibitors as part of the step-by-step process of drug design was established. The reference protein (PDB codes: 2CG9, 3PRY and 3HJC) crystal structures were used as templates to construct a human Hsp90β homologue. The Ramachandran plot showed that the validity of the human Hsp90 homologue was 98% of all residues, including the active site residues, which were in the favored region with 99.8% being in the allowed region. 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 (1), and this formed the basis of investigating and validating these theories by computational methods.

4. 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.

5. The RMSDs of all the nine complexes at the NTD did not exceed 2Å and the system was stabilized after 1000 ps MD simulation, indicating that the system was well equilibrated. On the other hand, the CTD domain showed 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 ≤8000 kcal/mol. Overall, it is observed that equilibration at NTD and CTD was achieved, although they happen at different time frames.

6. From the above results, 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 included: Leu43, Asn46, Lys53, Ile91, Asp97, Met93, Asn101, Ser108, Gly109, Phe133 and Thr179. Sufficient computational evidence has indicated 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.

6.2. Recommendations and Future Studies

This study enabled the identification of potential Michael acceptor-type compounds that form irreversible binding with Cathepsin B, and with the mechanism of inhibition of HIV-1 protease inhibitors on Hsp90 also being explored, all for cancer eradication. However, further computational investigation will be necessary to understand enzyme-substrate interactions and structural evolution towards substrate binding. Quantum Mechanics implemented in this study could provide better insights into the molecular interactions regarding the distribution of motion by defining the behavior of nuclei and electrons within the enzyme-inhibitor complexes. It could also help to determine the nature of enzyme-inhibitor bond formation in order to improve our understanding into their binding mode (*2, 3*). Other advanced computational approaches that would be useful include Residue Interaction Network (RIN) (*4*), Principal Component Analysis (PCA) (*5*), Substrate Envelope Analysis (SEA) (*6*), Quantitative Structure-Activity Relationships (QSAR) (*7*), and coarse grained molecular dynamics (*8*) as they provide a better understanding of the enzyme dynamics during the catalytic course. In addition, further experimental testing would be confirmatory to these findings.

References

1.   Peterson, L. B. (2012) Investigation of the Hsp90 C-terminal binding site, Novel inhibitors and isoform-dependent client proteins.

2.   Mucs, D., and Bryce, R. A. (2013) The application of quantum mechanics in structure-based drug design, *Expert. Opin. Drug Discov. 8*, 263-276.

3.   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.

4.   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.

5.   Maisuradze, G. G., Liwo, A., and Scheraga, H. A. (2009) Principal Component Analysis for Protein Folding Dynamics, *Journal of Molecular Biology 385*, 312-329.

6.   Shen, Y., Altman, M. D., Ali, A., Nalam, M. N. L., Cao, H., Rana, T. M., Schiffer, C. A., and Tidor, B. (2013) Testing the Substrate-Envelope Hypothesis with Designed Pairs of Compounds, *ACS Chem. Biol. 8*, 2433-2441.

7.   Hirst, J. D., McNeany, T. J., Howe, T., and Whitehead, L. (2002) Application of non-parametric regression to quantitative structure-activity relationships, *Bioorg. Med. Chem. 10*, 1037-1041.

8.   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 A1: Input files for docking the human cathepsin B complexes

receptor = rec.pdbqt

exhaustiveness = 8

center_x = 33.9

center_y = 34.1

center_z = 32.9

size_x = 34

size_y = 42

size_z = 48

Appendix A2: Input files for docking the modelled human Hsp90 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 A3: Input files for docking the modelled human Hsp90 complexes at CTD

receptor = rec.pdbqt

exhaustiveness = 8

center_x = -81.0

center_y = -54.0

center_z = 56.0

size_x = 16

size_y = 22

size_z = 28

Appendix B: Supplementary material for 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*]

Figure 6: Comparative RMSF plot for PIs at the human Hsp90 homologue NTD and CTD
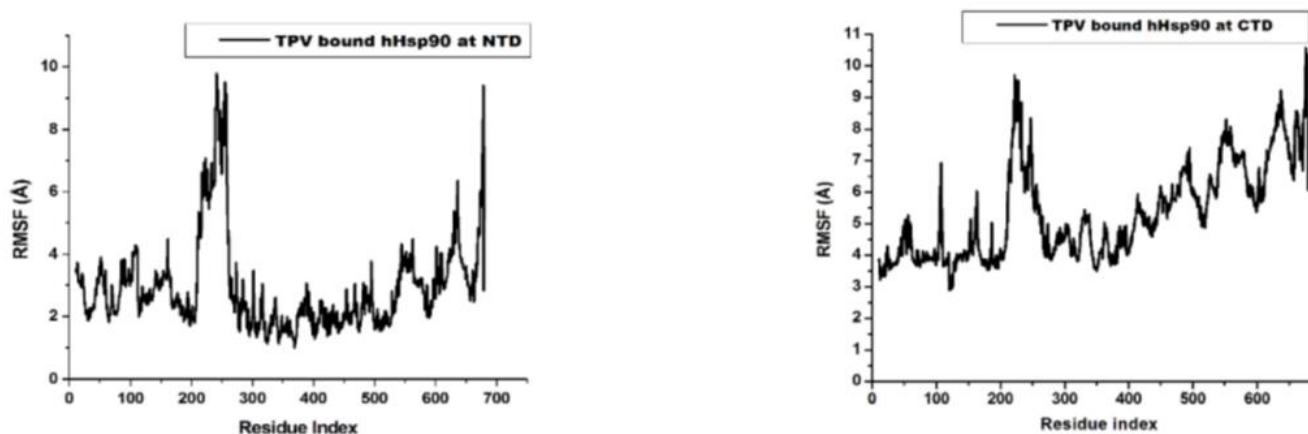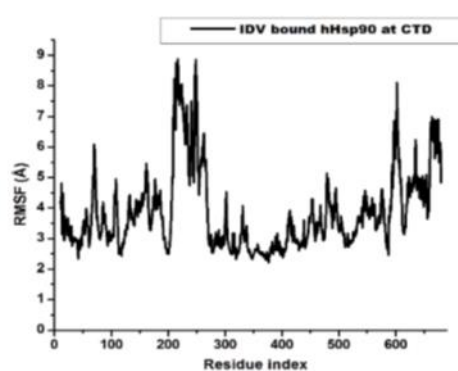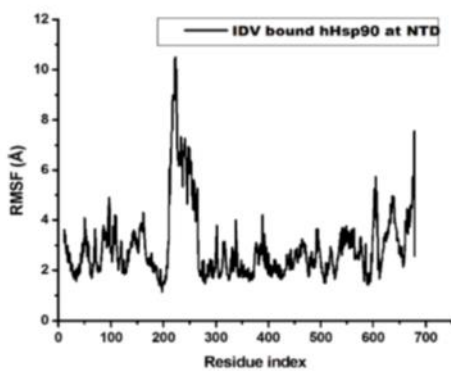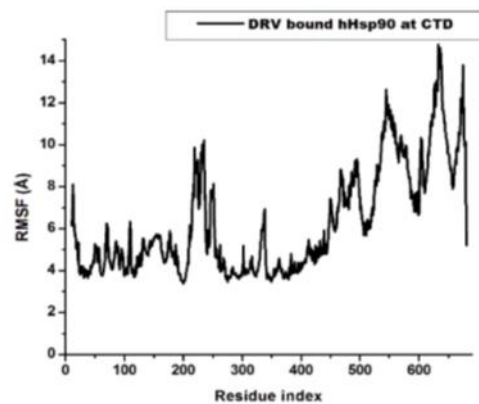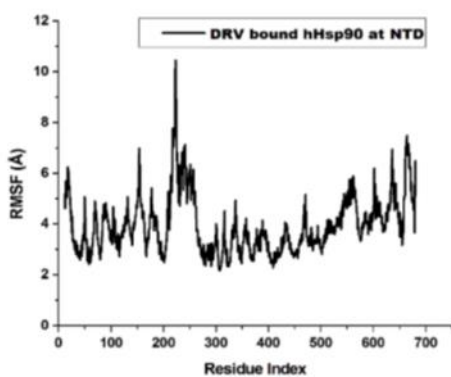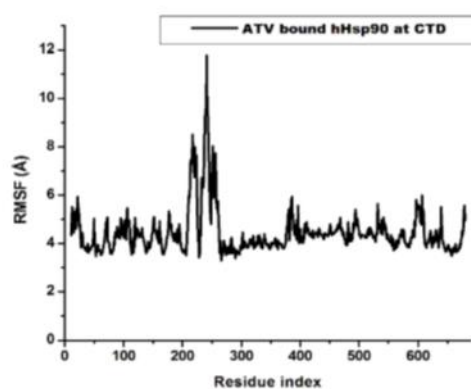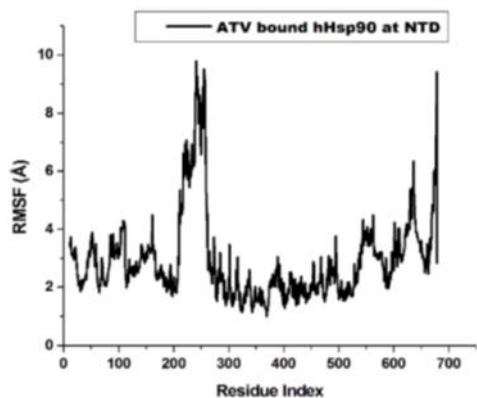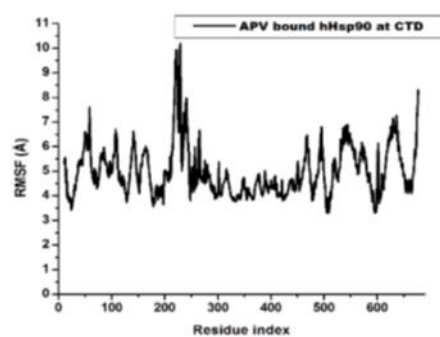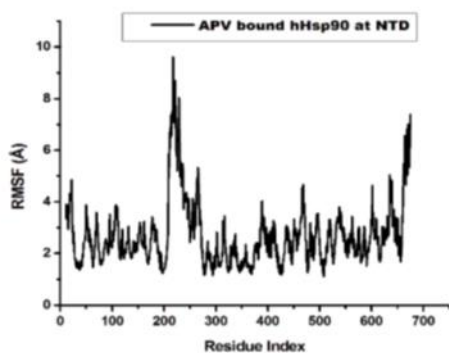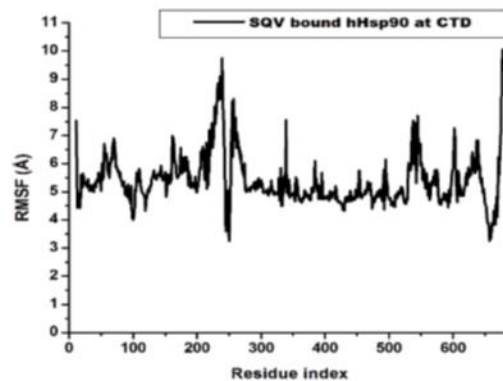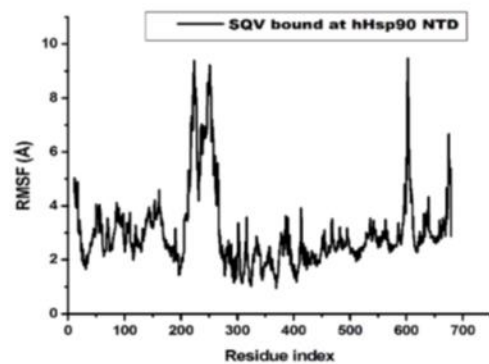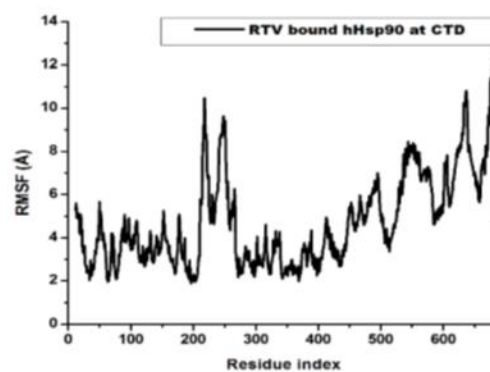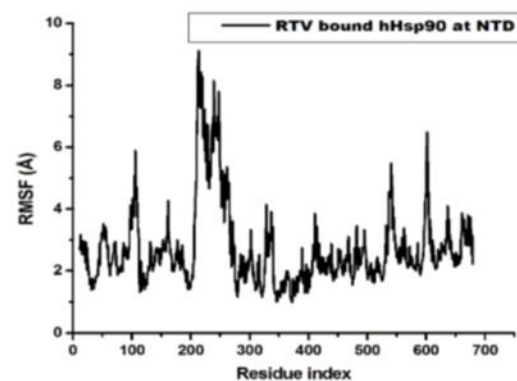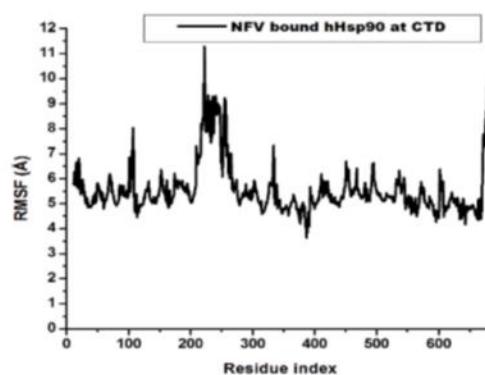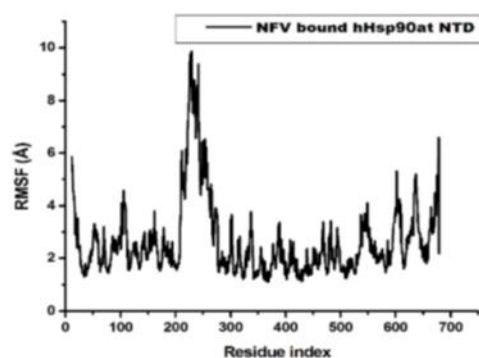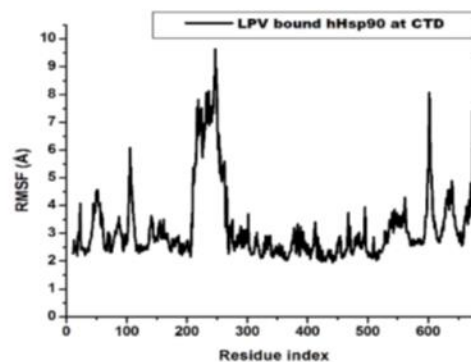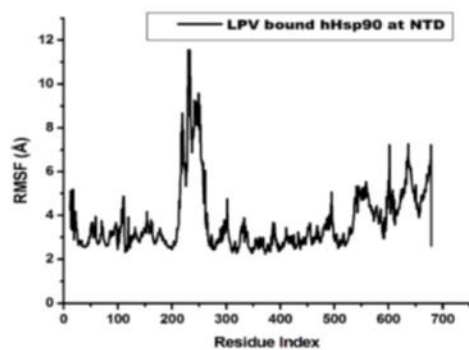
Figure 7(a): Comparative RMSD plot for PIs at the human Hsp90 homologue NTD and CTD
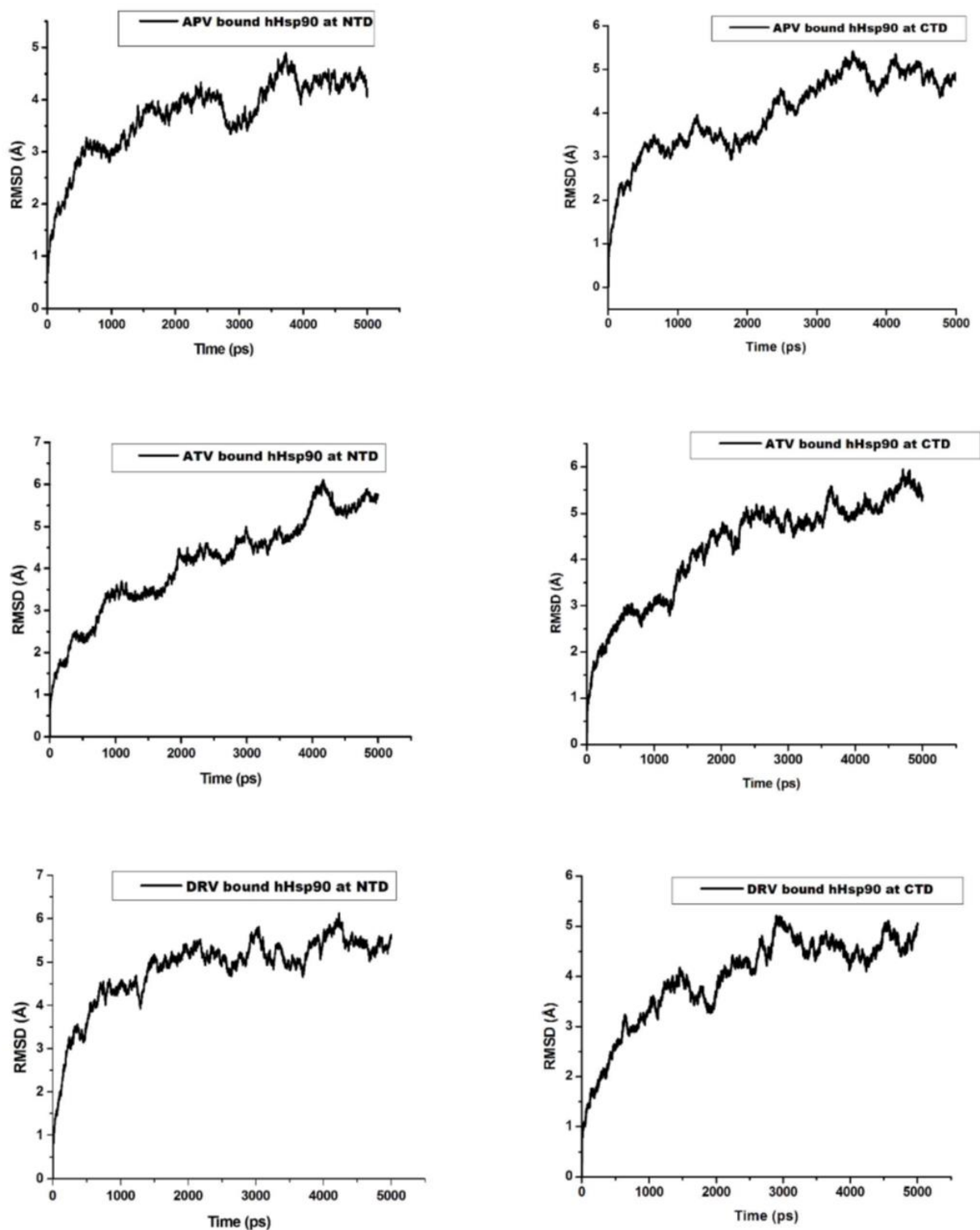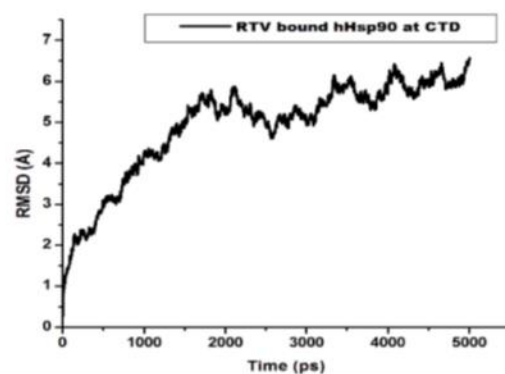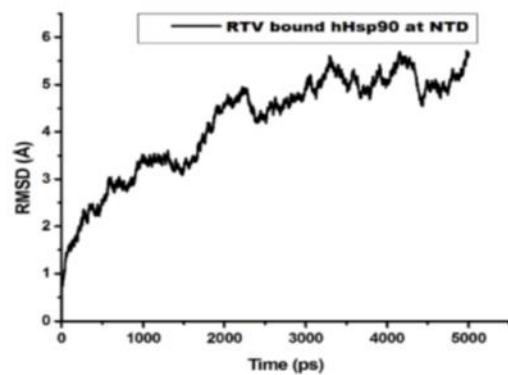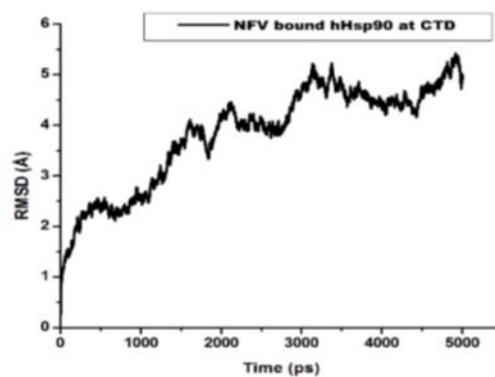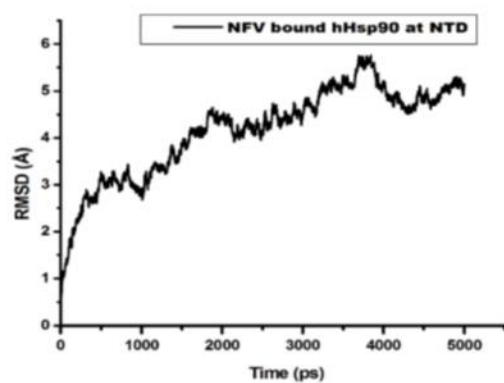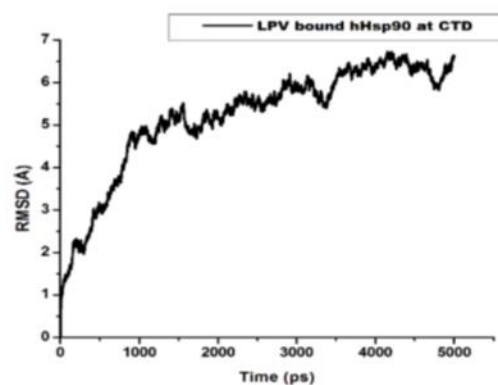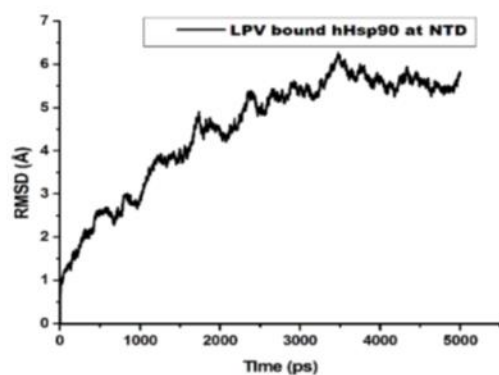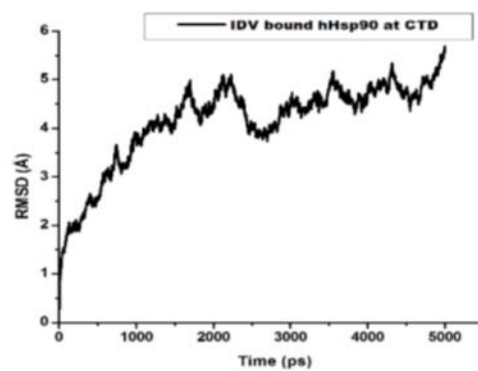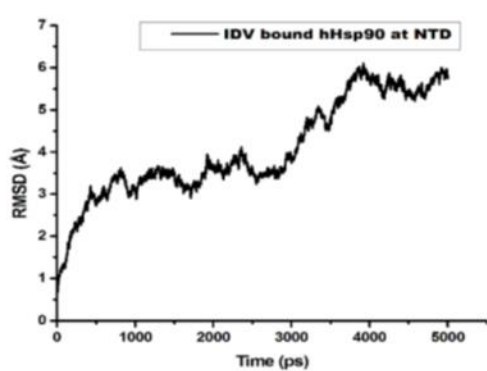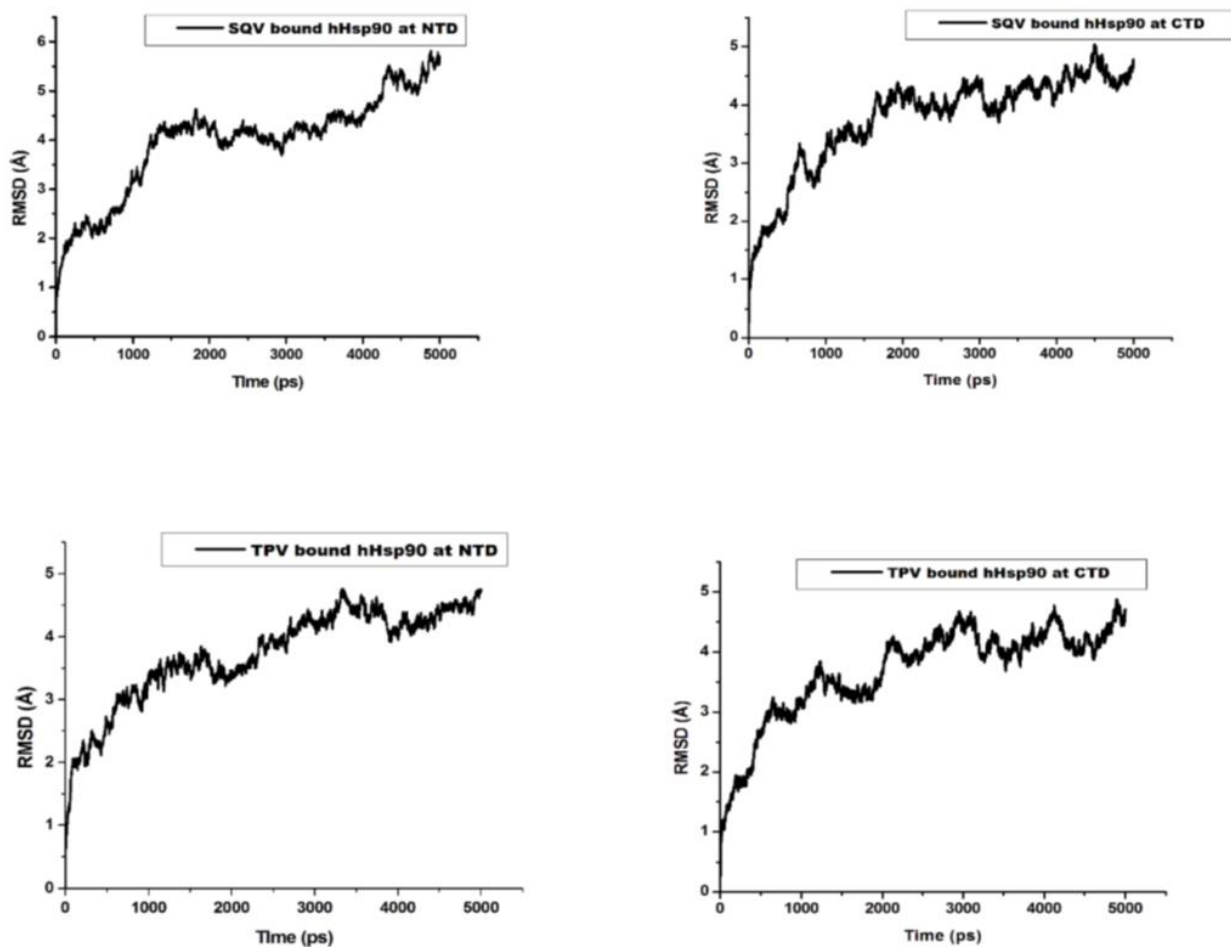
Figure 7(b): Comparative potential energy plot for PIs at the human Hsp90 homologue
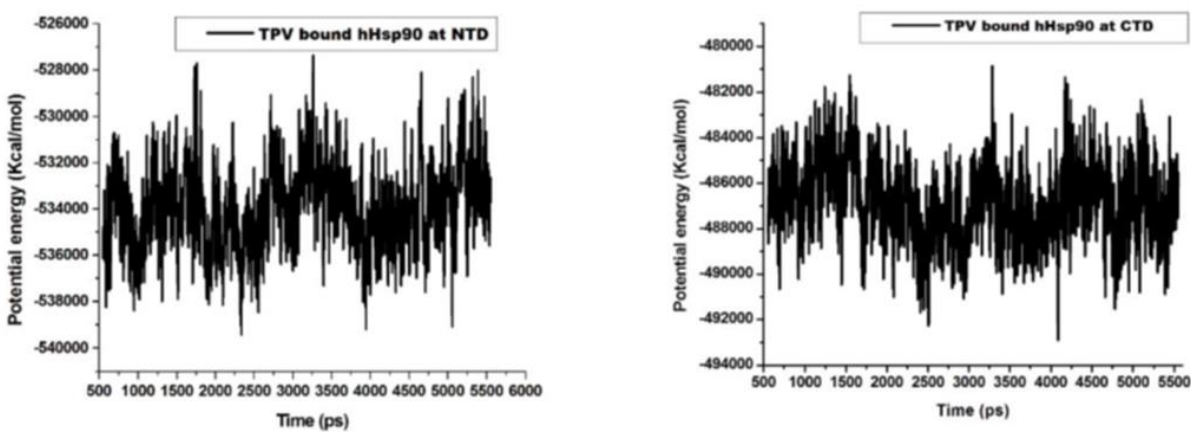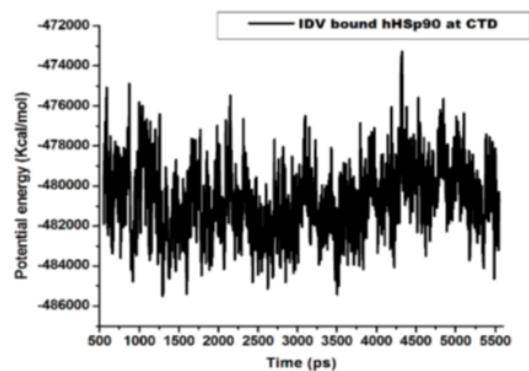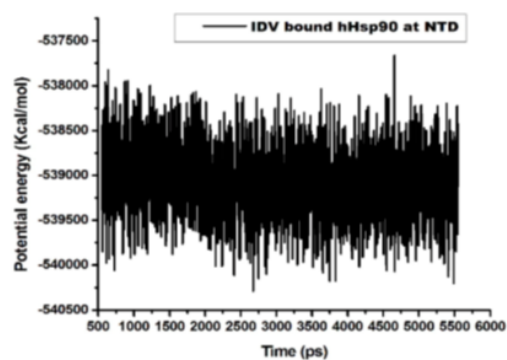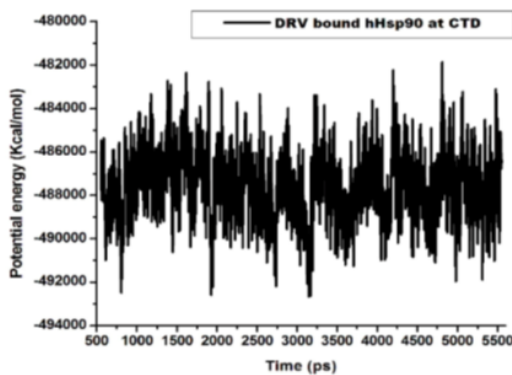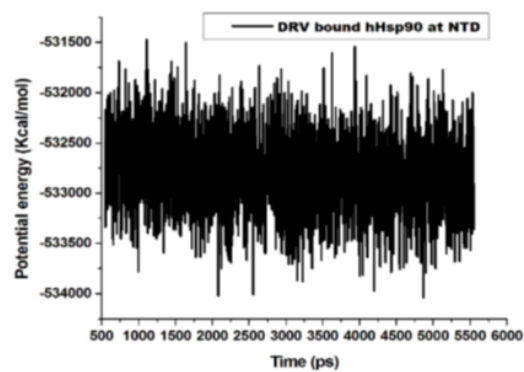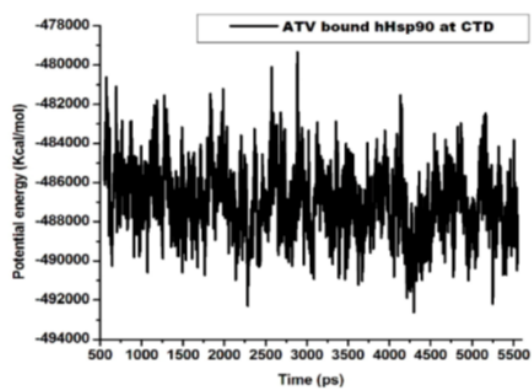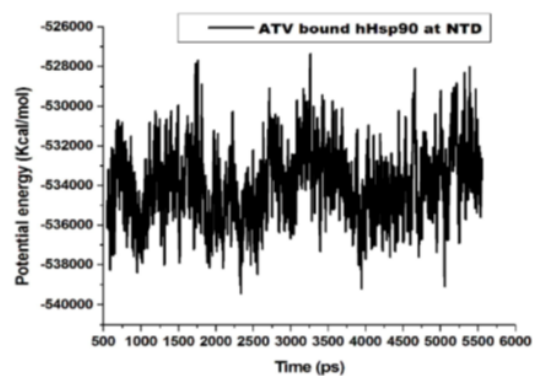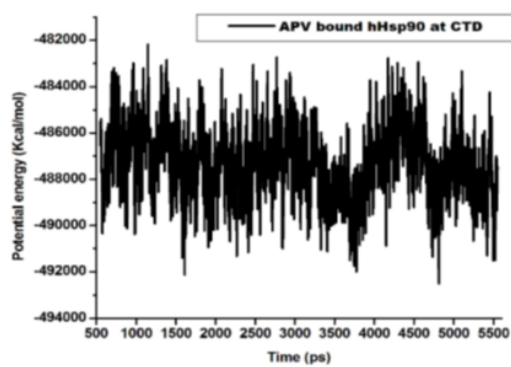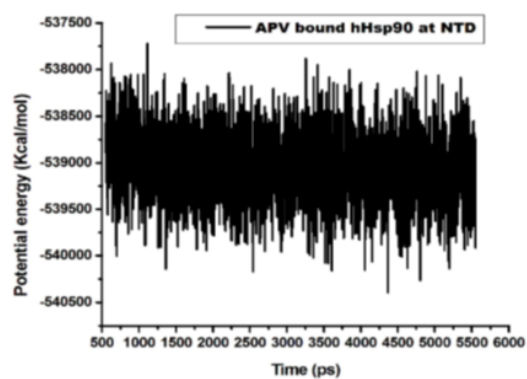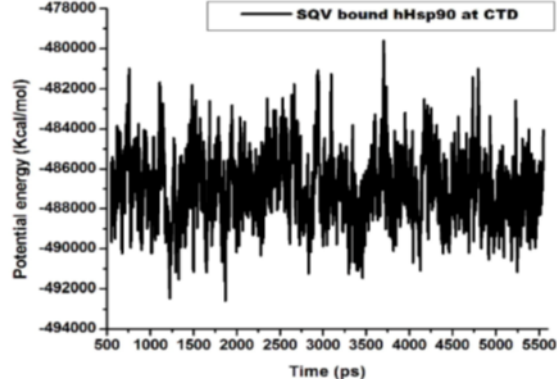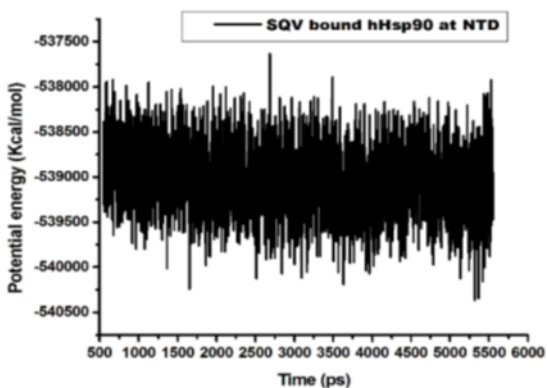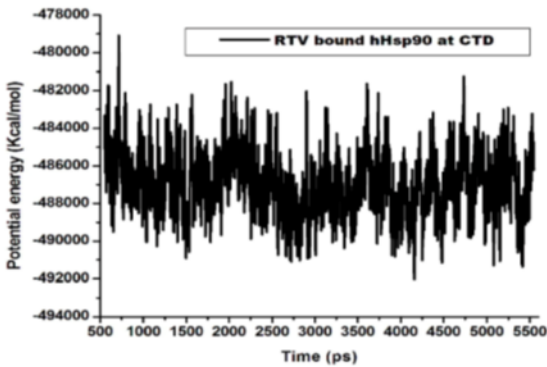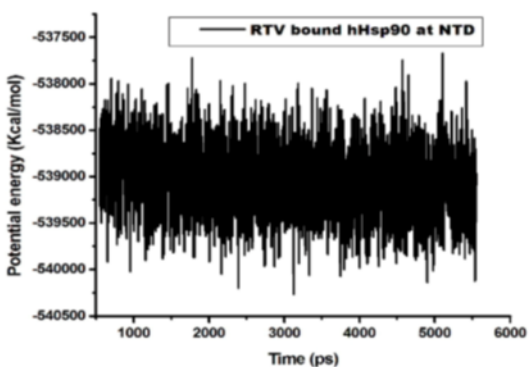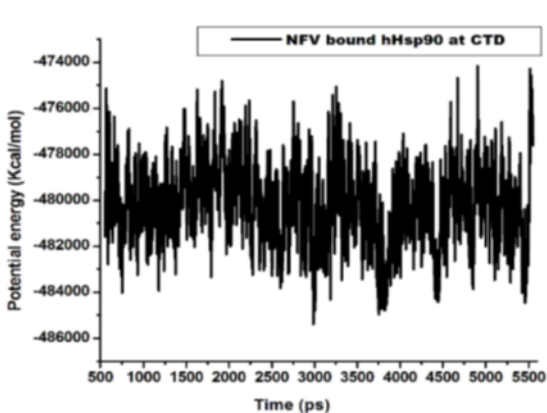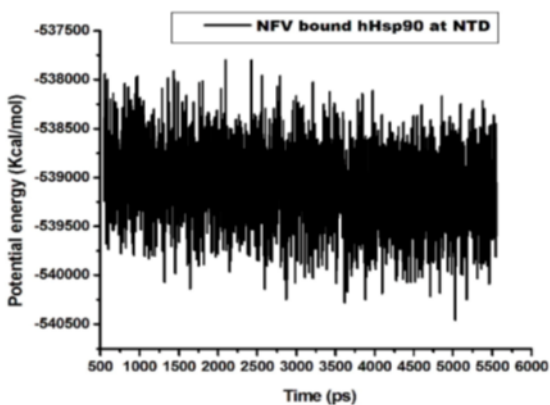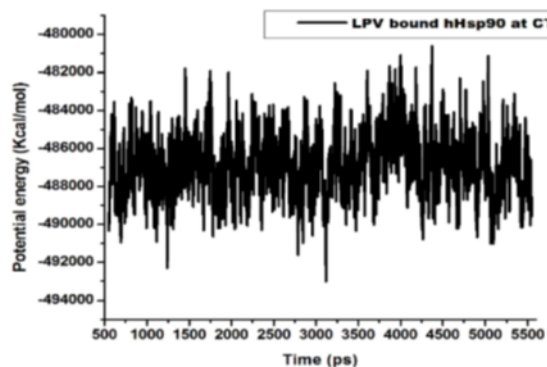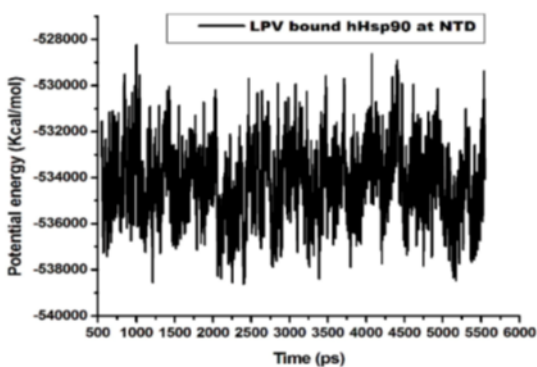
Supplementary Figure-S1: The multiple sequence alignment result from CLUSTAW

```
CLUSTAL 2.1 multiple sequence alignment


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


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


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


3PRY_A|PDBID|CHAIN|SEQUENCE        ----------------------------------MKTKPIWTRNPDDIT 15
sp|P08238|HS90B_HUMAN              PKIEDVGSDEEDDSGKDKKKKTKKIKEKYIDQEELNKTKPIWTRNPDDIT 297
2CG9_B|PDBID|CHAIN|SEQUENCE        PKLEEVDEEEE------KKPKTKKVKEEVQEIEELNKTKPLWTRNPSDIT 285
3HJC_A|PDBID|CHAIN|SEQUENCE        --------------------MHHHHHHSSGRENLYFQGHKPLWTRDPKDVT 31
                                                      **:***:*.*:*

3PRY_A|PDBID|CHAIN|SEQUENCE        QEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFENK 65
sp|P08238|HS90B_HUMAN              QEEYGEFYKSLTNDWEDHLAVKHFSVEGQLEFRALLFIPRRAPFDLFENK 347
2CG9_B|PDBID|CHAIN|SEQUENCE        QEEYNAFYKSISNDWEDPLYVKHFSVEGQLEFRAILFIPKRAPFDLFESK 335
3HJC_A|PDBID|CHAIN|SEQUENCE        KEEYAAFYKAISNDWEDPAATKHFSVEGQLEFRSIMFVPKRAPFDMFEPN 81
                                   :***  ***:;:*****  .*************;::*:*:*****:** :
```

115

```
3PRY_A|PDBID|CHAIN|SEQUENCE    KKKNNIKLYVRRVFIMDSCDELIPEYLNFIRGVVDSEDLPLNISREMLQQ 115
sp|P08238|HS90B_HUMAN          KKKNNIKLYVRRVFIMDSCDELIPEYLNFIRGVVDSEDLPLNISREMLQQ 397
2CG9_B|PDBID|CHAIN|SEQUENCE    KKKNNIKLYVRRVFITDEAEDLIPEWLSFVKGVVDSEDLPLNLSREMLQQ 385
3HJC_A|PDBID|CHAIN|SEQUENCE    KKRNNIKLYVRRVFIMDNCEDLCPDWLGFVKGVVDSEDLPLNISRENLQQ 131
                               **:************ *..::* *::*.*::************:*** ***

3PRY_A|PDBID|CHAIN|SEQUENCE    SKILKVIRKNIVKKCLELFSELAEDKENYKKFYEAFSKNLKLGIHEDSTN 165
sp|P08238|HS90B_HUMAN          SKILKVIRKNIVKKCLELFSELAEDKENYKKFYEAFSKNLKLGIHEDSTN 447
2CG9_B|PDBID|CHAIN|SEQUENCE    NKIMKVIRKNIVKKLIEAFNEIAEDSEQFEKFYSAFSKNIKLGVHEDTQN 435
3HJC_A|PDBID|CHAIN|SEQUENCE    NKILKVIRKNIVKKCLEMFDEVAENKEDYKQFYEQFGKNIKLGIHEDTAN 181
                               .**:********** :* *.*:**:.*:::** *.**:***:***: *

3PRY_A|PDBID|CHAIN|SEQUENCE    RRRLSELLRYHTSQSGDEMTSLSEYVSRMKETQKSIYYITGESKEQVANS 215
sp|P08238|HS90B_HUMAN          RRRLSELLRYHTSQSGDEMTSLSEYVSRMKETQKSIYYITGESKEQVANS 497
2CG9_B|PDBID|CHAIN|SEQUENCE    RAALAKLLRYNSTKSVDELTSLTDYVTRMPEHQKNIYYITGESLKAVEKS 485
3HJC_A|PDBID|CHAIN|SEQUENCE    RKKLMELLRFYSTESGEEMTTLKDYVTRMKAGQKSIYYITGDSKKKLETS 231
                               *   * :***: :::* :*:*:*.:**:**   **.******:* : : .*

3PRY_A|PDBID|CHAIN|SEQUENCE    AFVERVRKRGFEVVYMTEPIDEYCVQQLKEFDGKSLVSVTKEGLELAEN- 264
sp|P08238|HS90B_HUMAN          AFVERVRKRGFEVVYMTEPIDEYCVQQLKEFDGKSLVSVTKEGLELPEDE 547
2CG9_B|PDBID|CHAIN|SEQUENCE    PFLDALKAKNFEVLFLTDPIDEYAFTQLKEFEGKTLVDITKD-FELEETD 534
3HJC_A|PDBID|CHAIN|SEQUENCE    PFIEQARRRGLEVLFMTEPIDEYVMQQVKDFEDKKFACLTKEGVHFEESE 281
                               .*::   : :.:**::;*:*****  . *:*:*:.*.:. :**: ..: *

3PRY_A|PDBID|CHAIN|SEQUENCE    -----------LYFQ------------------------------------ 268
sp|P08238|HS90B_HUMAN          EEKKKMEESKAKFENLCKLMKEILDKKVEKVTISNRLVSSPCCIVTSTYG 597
2CG9_B|PDBID|CHAIN|SEQUENCE    EEKAEREKEIKEYEPLTKALKEILGDQVEKVVVSYKLLDAPAAIRTGQFG 584
3HJC_A|PDBID|CHAIN|SEQUENCE    EEKQQREEEKAACEKLCKTMKEVLGDKVEKVIVSERLSTSPCILVTSEFG 331
                                              :

3PRY_A|PDBID|CHAIN|SEQUENCE    -------------------------------------------------
sp|P08238|HS90B_HUMAN          WTANMERIMKAQALRDNSTMGYMMAKKHLEINPDHPIVETLRQKA-EADK 646
2CG9_B|PDBID|CHAIN|SEQUENCE    WSANMERIMKAQALRDSSMSSYMSSKKTFEISPKSPIIKELKKRVDEGGA 634
3HJC_A|PDBID|CHAIN|SEQUENCE    WSAHMEQIMRNQALRDSSMAQYMMSKKTMELNPRHPIIKELRRRV-GADE 380

3PRY_A|PDBID|CHAIN|SEQUENCE    -------------------------------------------------
sp|P08238|HS90B_HUMAN          NDKAVKDLVVLLFETALLSSGFSLEDPQTHSNRIYRMIKLGLGIDEDEVA 696
2CG9_B|PDBID|CHAIN|SEQUENCE    QDKTVKDLTKLLYETALLTSGFSLDEPTSFASRINRLISLGLN------- 677
3HJC_A|PDBID|CHAIN|SEQUENCE    NDKAVKDLVFLLFDTSLLTSGFQLEDPTGYAERINRMIKLGLSLDEEEEE 430

3PRY_A|PDBID|CHAIN|SEQUENCE    ------------------------------
sp|P08238|HS90B_HUMAN          AEEPNAAVPDEIPPLEGDEDASRMEEVD 724
2CG9_B|PDBID|CHAIN|SEQUENCE    ------------------------------
3HJC_A|PDBID|CHAIN|SEQUENCE    AAEAPVAETAPAEV--------------- 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-S2: Ramachandran plot for the human Hsp90 homologue
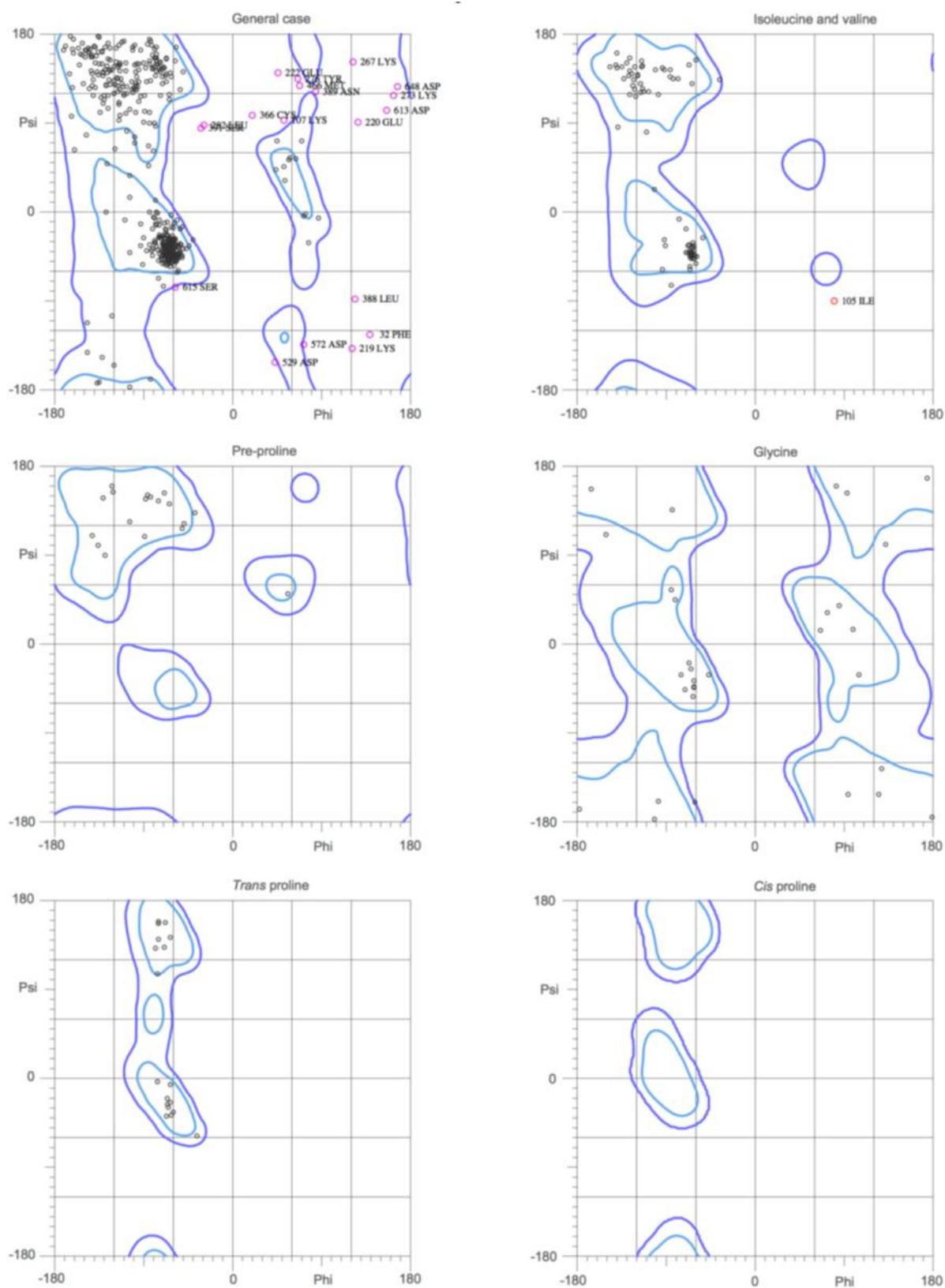
91.7% (620/676) of all residues were in favored (98%) regions.
97.0% (656/676) of all residues were in allowed (>99.8%) regions.

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-S3: Ligand-enzyme interaction at the human Hsp90 homologue NTD and CTD

APV at hHsp90 NTD
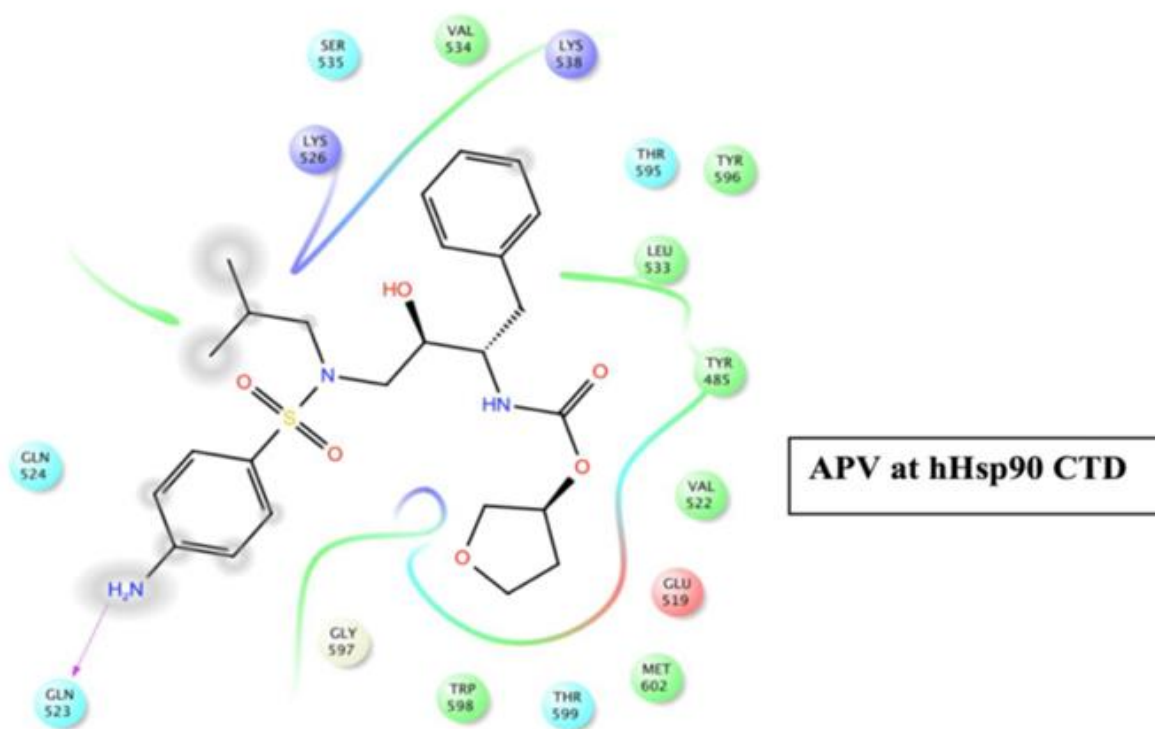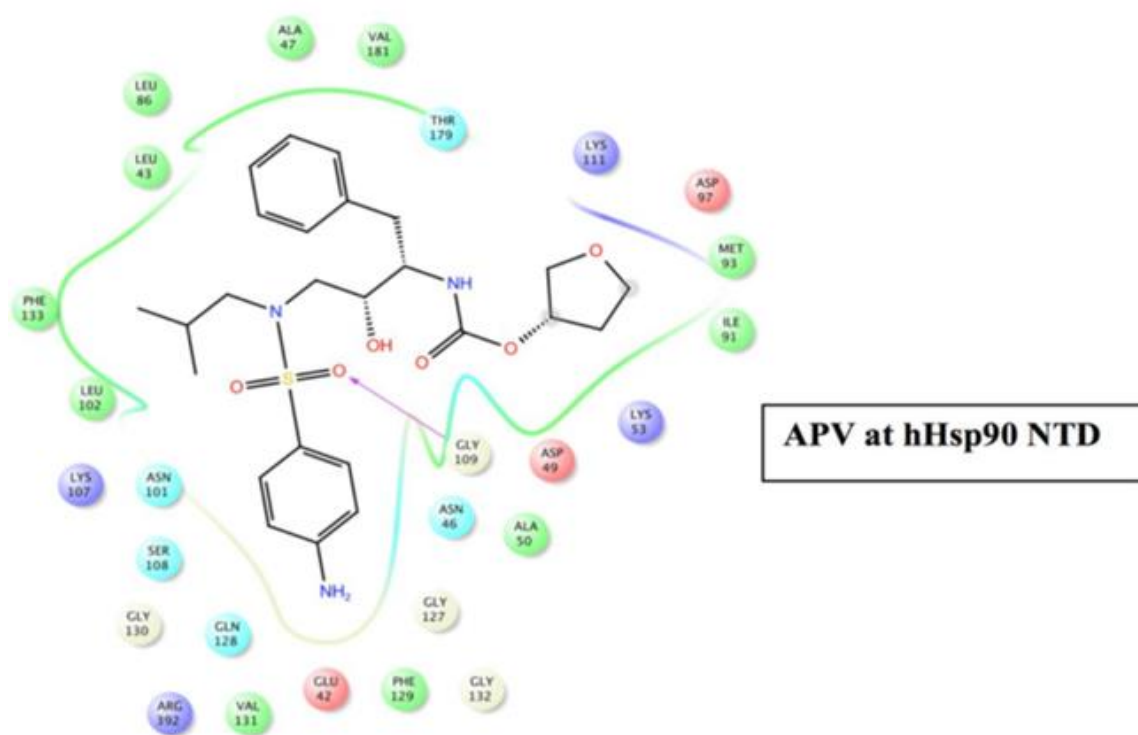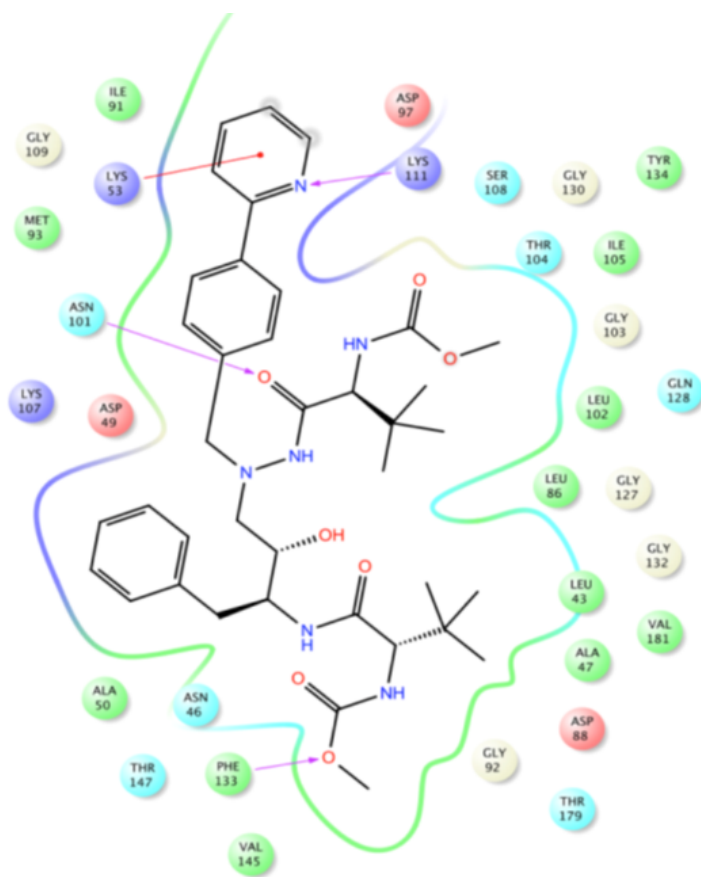
APV at hHsp90 CTD
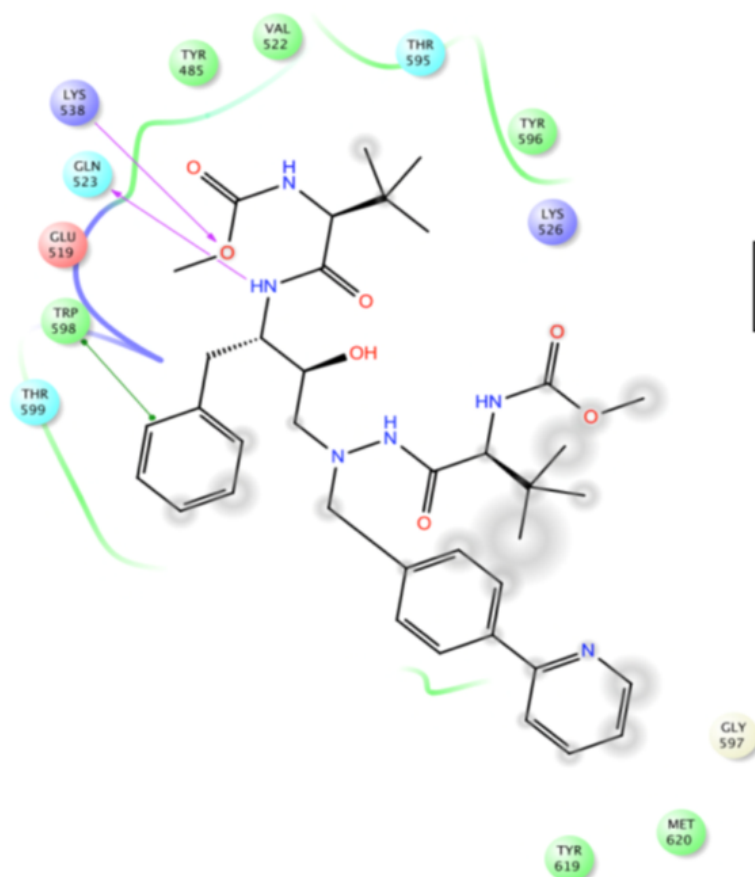
ATV at hHsp90 NTD

ATV at hHsp90 CTD
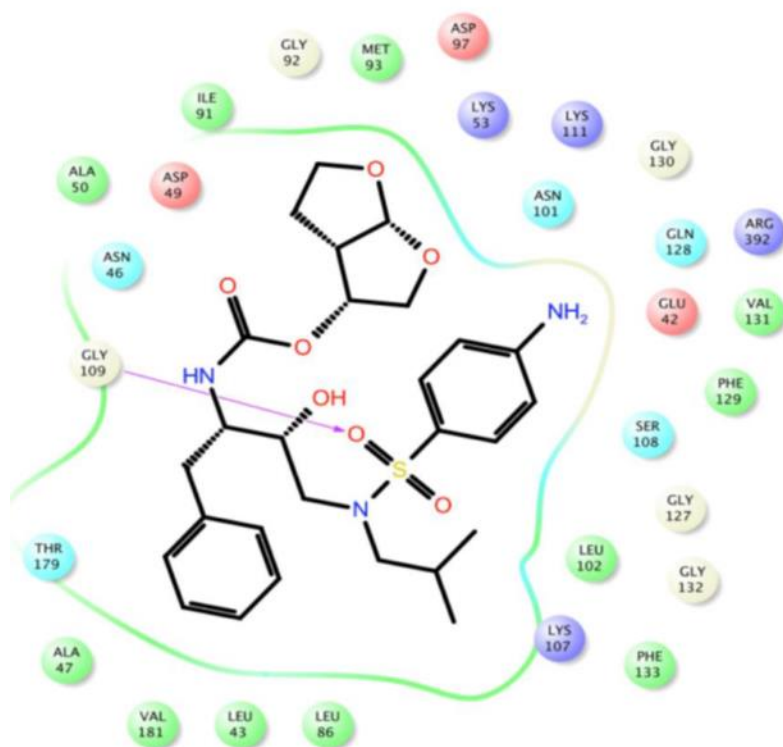
DRV at hHsp90 NTD
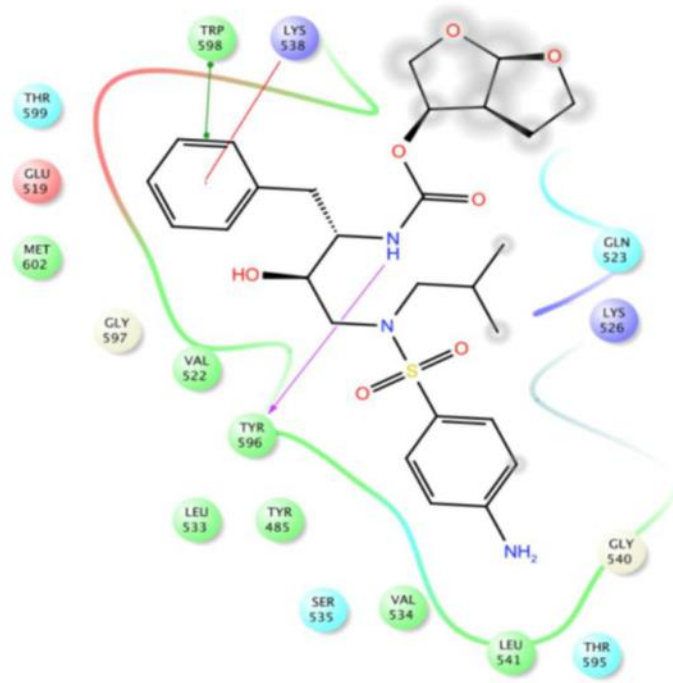
DRV at hHsp90 CTD
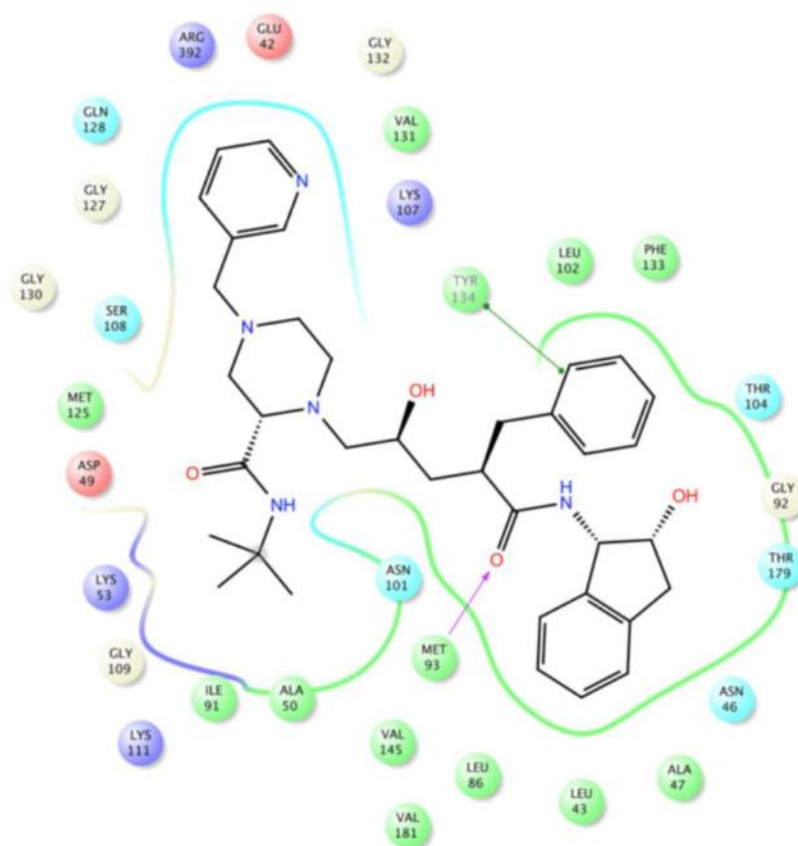
IDV at hHsp90 NTD
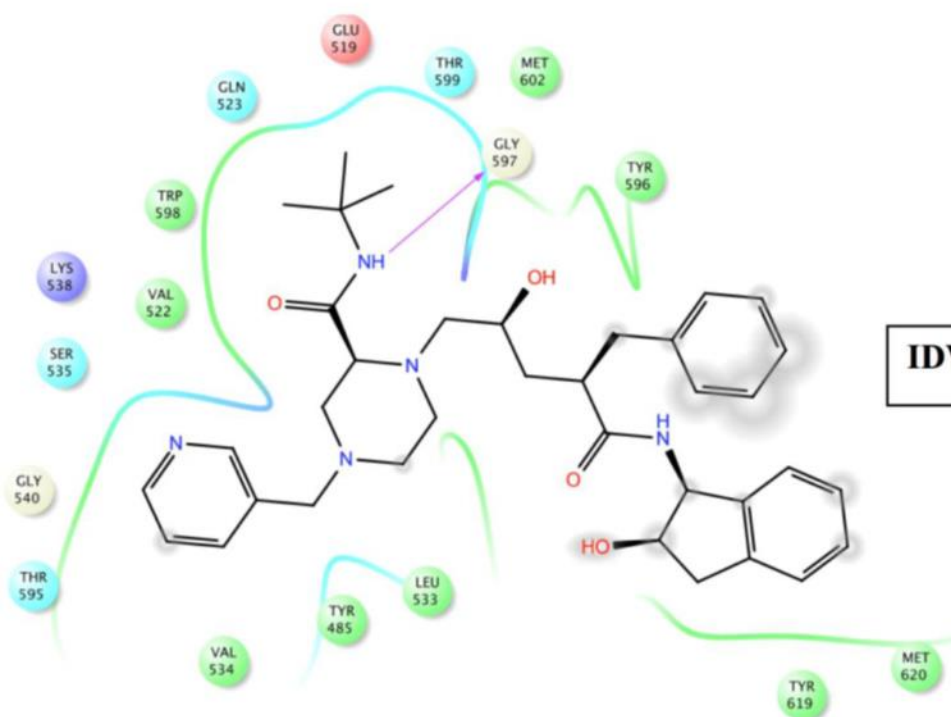
IDV at hHsp90 CTD
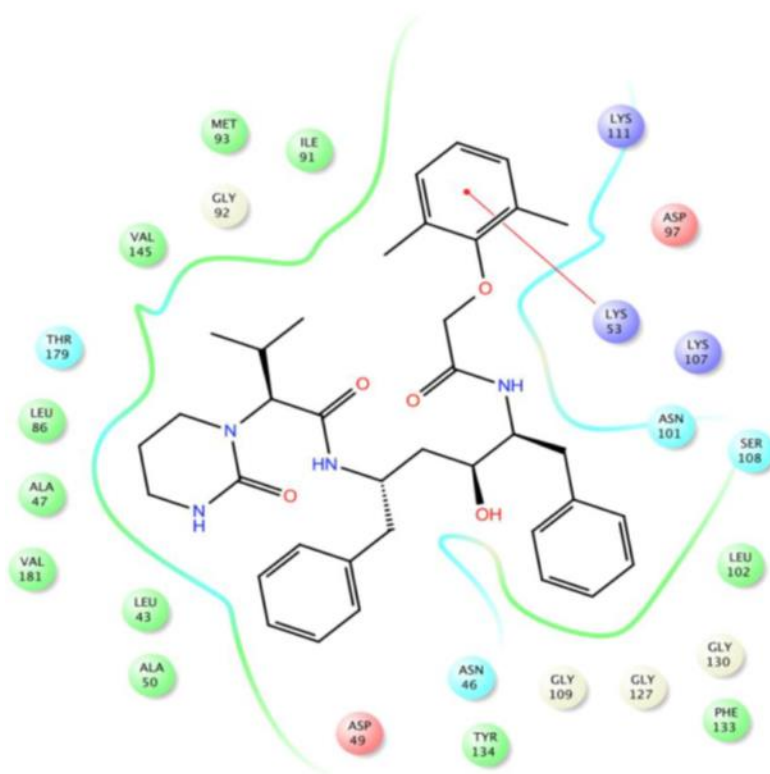
LPV at hHsp90 NTD

LPV at hHsp90 CTD
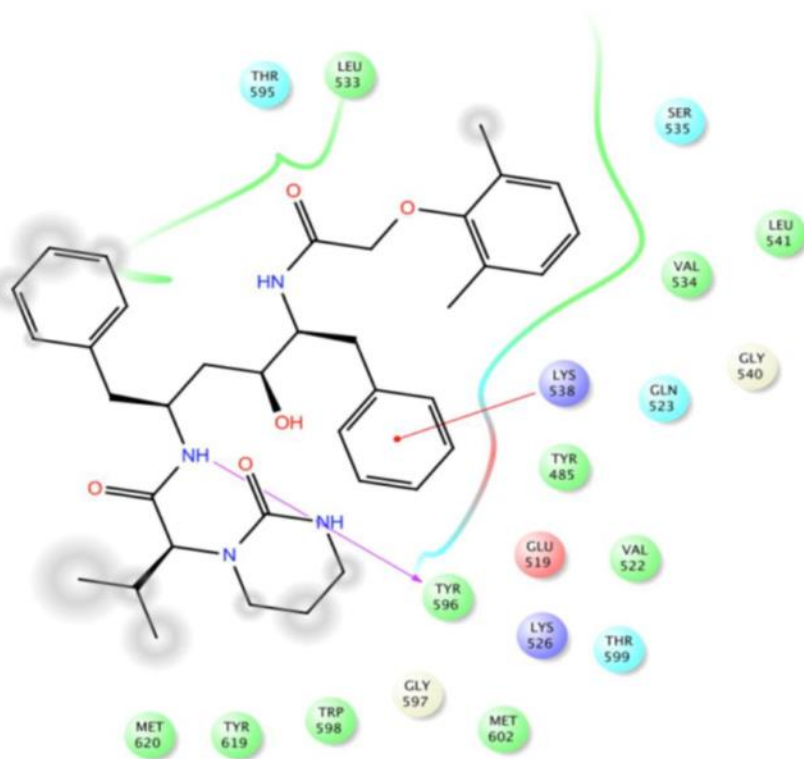
NFV at hHsp90 NTD

NFV at hHsp90 at CTD
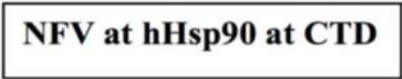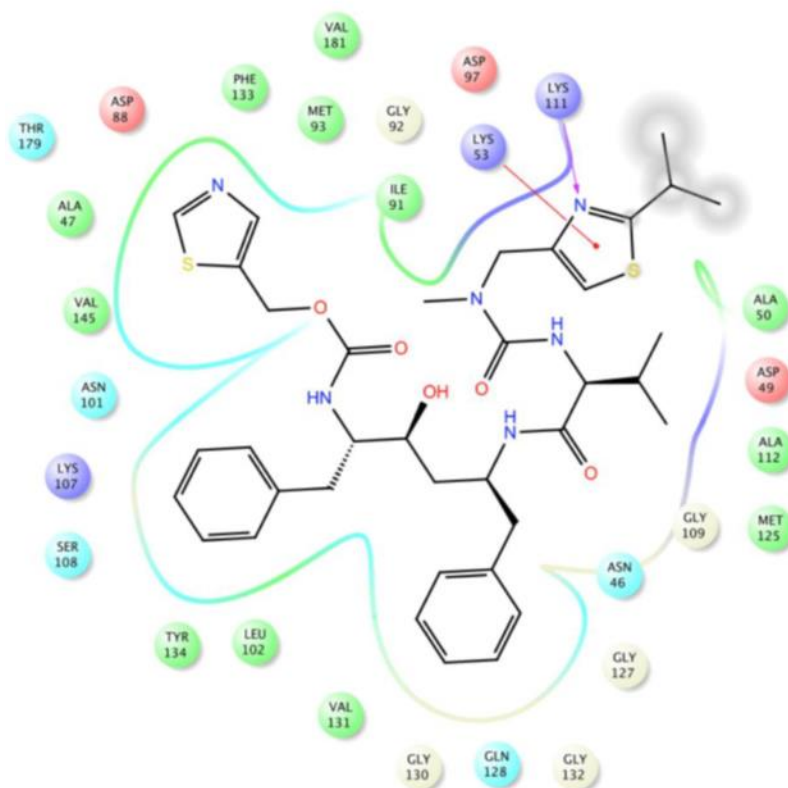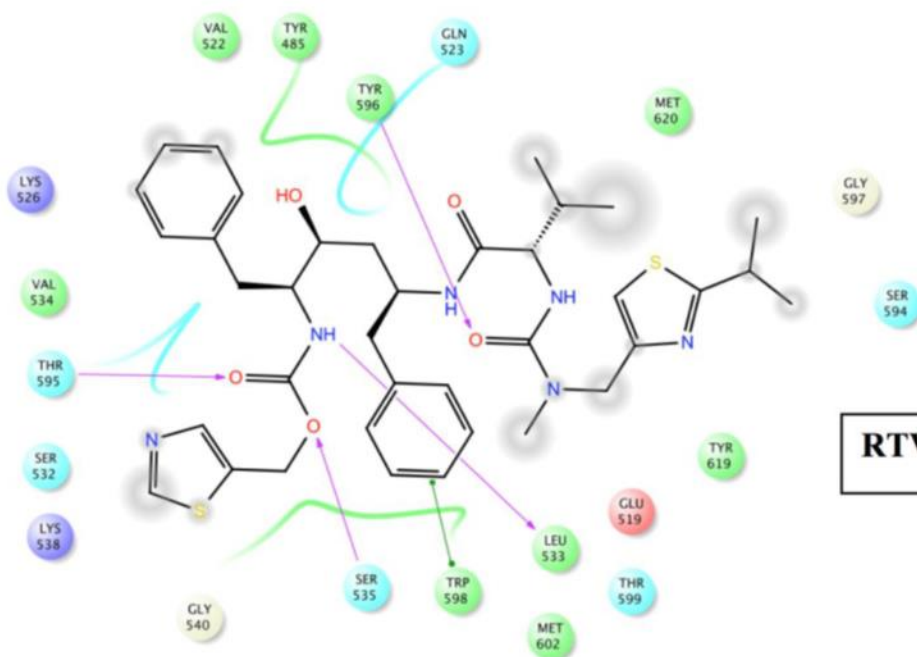
RTV at hHsp90 NTD

RTV at hHsp90 CTD
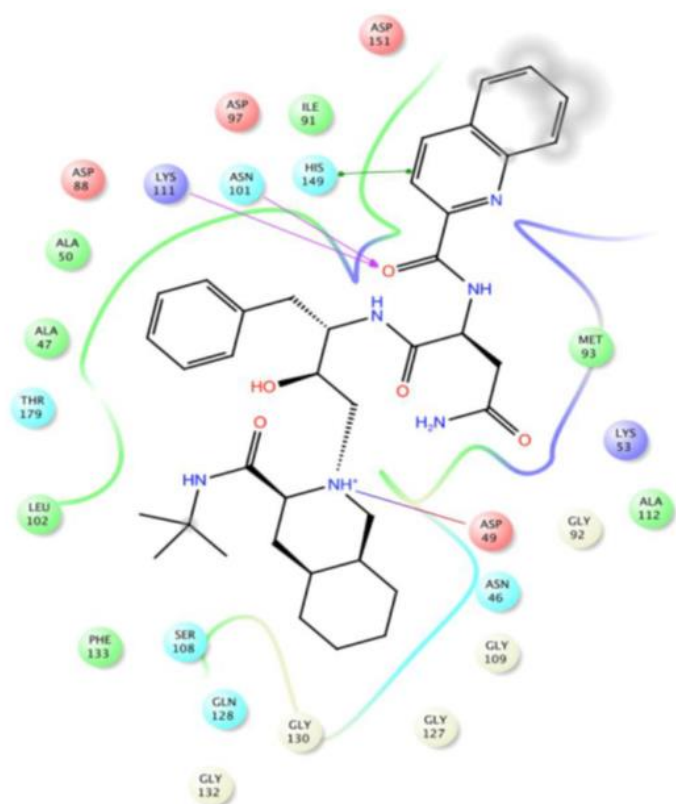
SQV at hHsp90 NTD

SQV at hHsp90 CTD

TPV at hHsp90 NTD
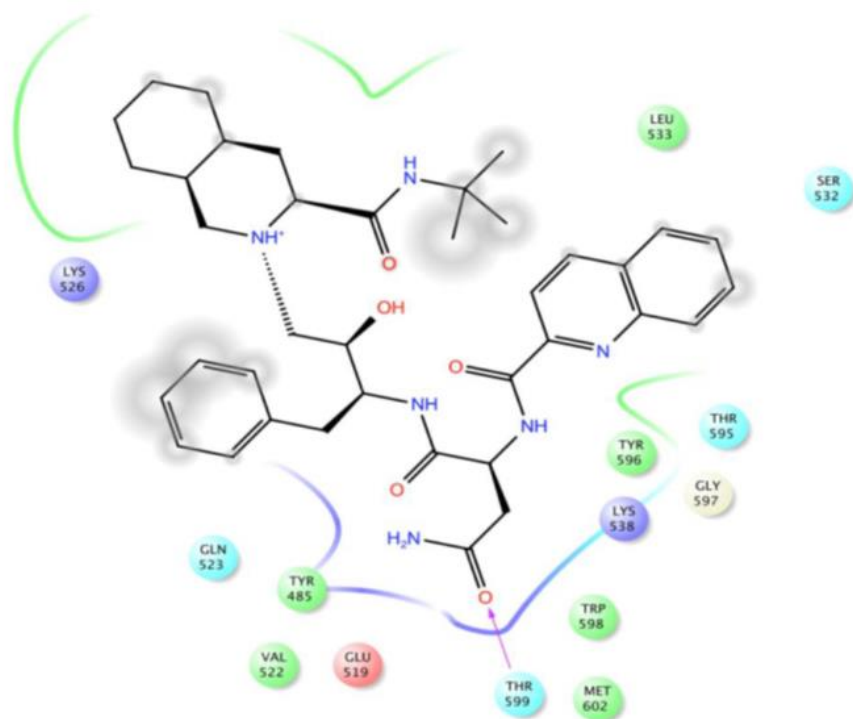
TPV at hHsp90 CTD
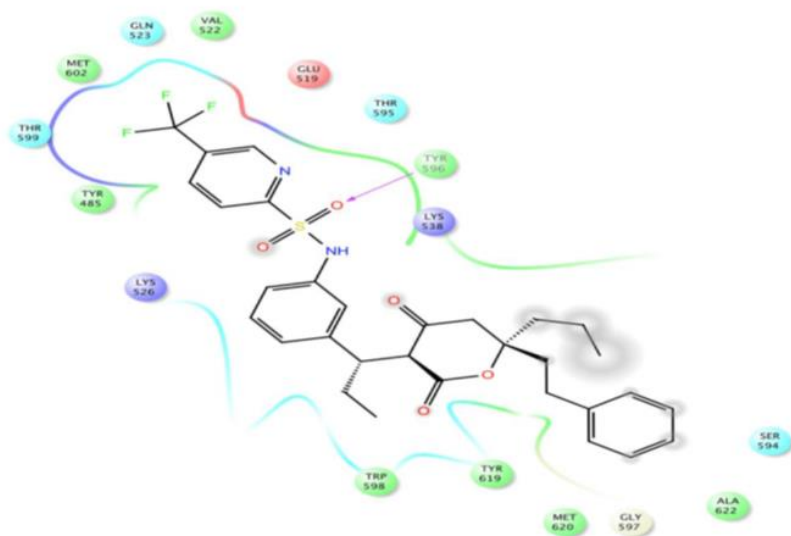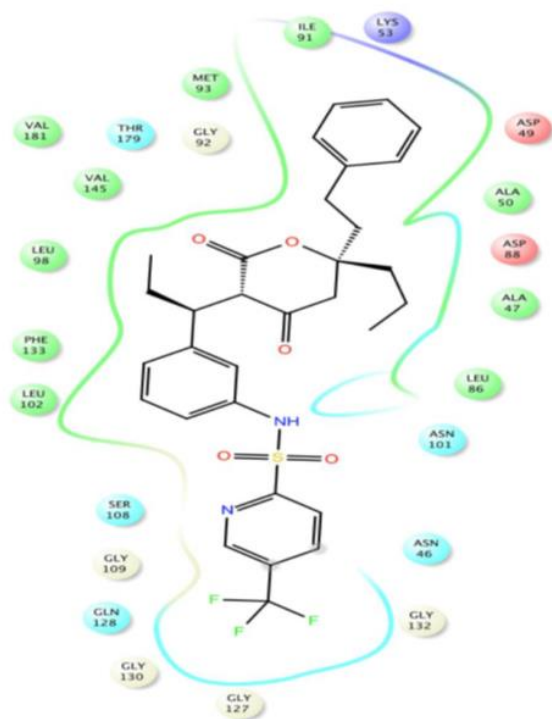
The legend for the ligand-enzyme interaction generated from Ligplot

Charged (negative)
Charged (positive)
Polar
Hydrophobic
Glycine
Metal

H₂O Water
Hydration site
Displaced hydration site
π–π stacking
π–cation
H–bond (backbone)

H–bond (side chain)
Metal coordination
Salt Bridge
Solvent exposure

Supplementary Figure-S4: Comparative per-residue decomposition energy at the human Hsp90 NTD and CTD
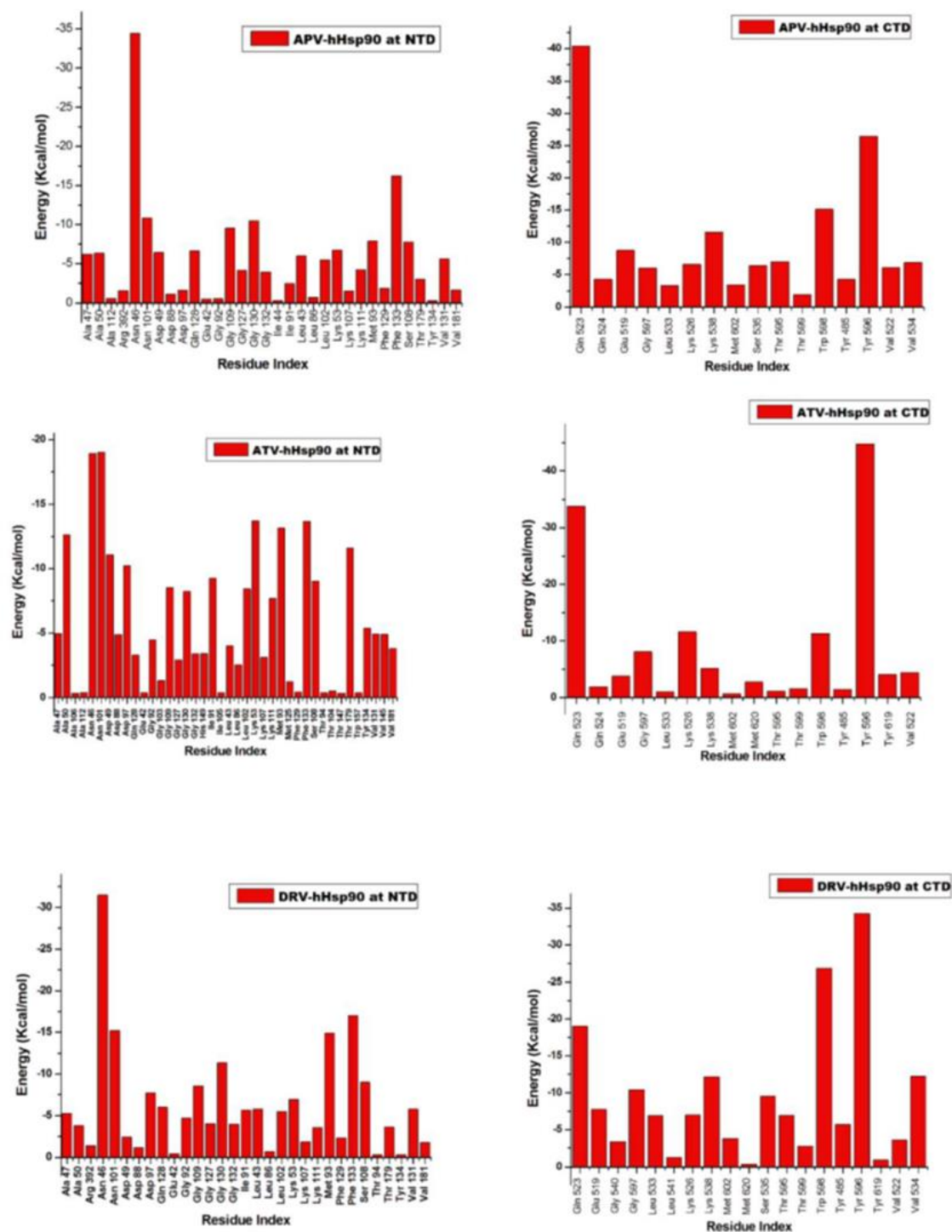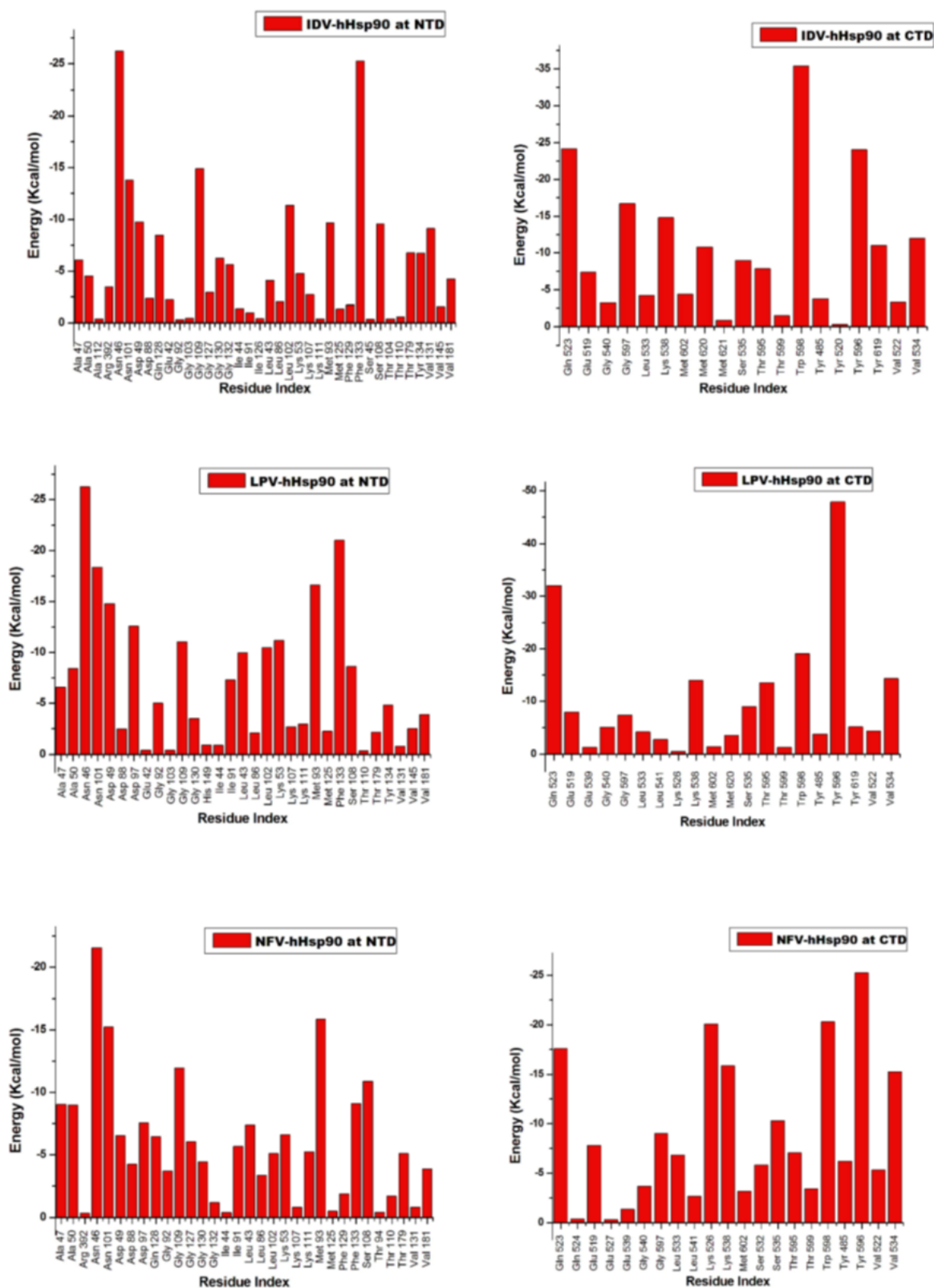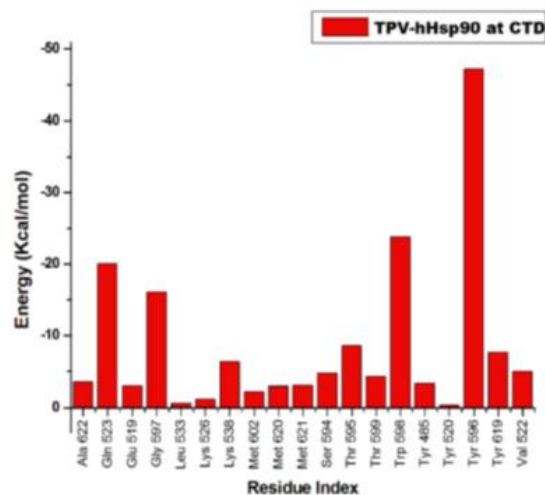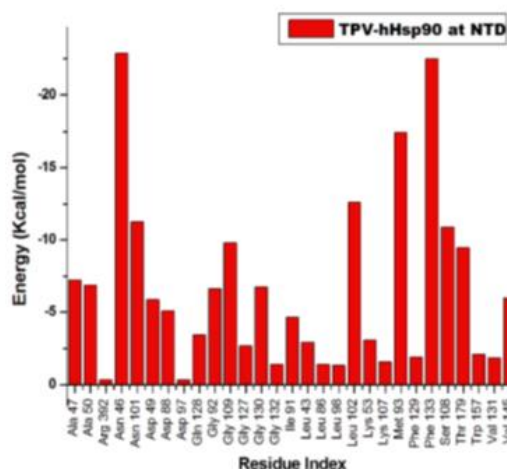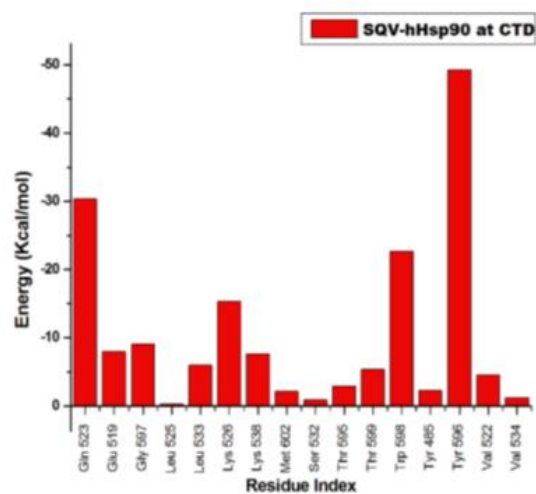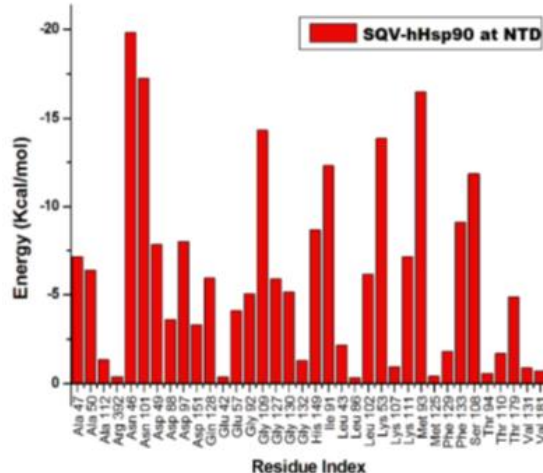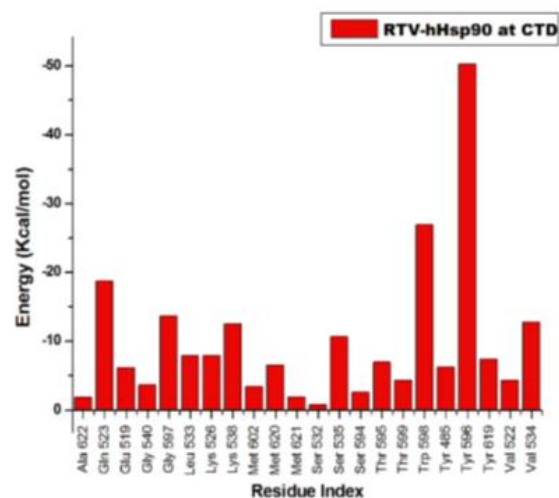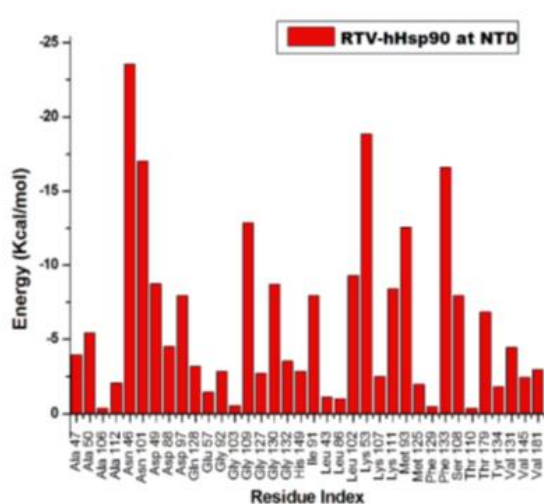
Appendix C. PDF version of the published paper