DIVERSE COMPUTATIONAL TOOLS TOWARDS THE UNDERSTNADING OF HIV TARGETS AND DESIGN OF POTENTIAL DRUG CANDIDATES

2014

SURI MOONSAMY

208508935

DIVERSE COMPUTATIONAL TOOLS TOWARDS THE UNDERSTNADING OF HIV TARGETS AND DESIGN OF POTENTIAL DRUG CANDIDATES

2014

SURI MOONSAMY

208508935

A thesis submitted to the School of Pharmacy and Pharmacology, **College of Health Sciences**, **University of KwaZulu-Natal**, **Westville**, for the **Degree of Doctor of Philosophy (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 **Miss Suri Moonsamy.**

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

Supervisor:

Signed: ----------------------   Name: **Dr. Mahmoud Soliman**   Date: ---------------------

ABSTRACT

HIV/AIDS still remains to be a challenging epidemic infecting millions of individuals worldwide. The morbidity and mortality rates of HIV-infected patients has been well documented over the years. Despite on-going HIV/AIDS research and access to antiretroviral therapy, to date still no cure exists for this deliberating disease.

In recent years, computational approaches have emerged as close counterparts to experiments in modern drug discovery process and in understanding complex biological phenomena. An array of *in-silico* computational techniques were implemented ranging from molecular dynamic (MD) simulations, *de-novo* design, hybrid structure-based and pharmacophore-based virtual screening, quantitative structure-activity relationship (QSAR), homology modeling, principle component analysis (PCA), residue interaction network analysis (RIN), substrate envelope analysis (SEA), to molecular mechanics and quantum mechanics.

The first report (Chapter 4), demonstrated a unique strategy for developing dual acting inhibitors against HIV-1 protease (PR) and reverse transcriptase (RT). The designed targets exhibited binding affinities and dual inhibiting activity comparable to, and in some cases better than, known active reference drugs.

The second study (Chapter 5), reported the activity of flexible hydroquinone-based compounds as non-nucleoside reverse transcriptase inhibitors (NNRTIs), as proposed by Bruccoleri, where no experimental or computational work supported his proposal. Results concluded that the novel flexible hydroquinone-based compounds showed improved binding affinity as compared to FDA-approved prototype drugs and more specifically potent potential mutant-resistant NNRT inhibitor activity.

The third report (Chapter 6), explored the activity of novel CCR5 antagonists as potential HIV-1 entry inhibitors. Ten scaffolds were identified as novel CCR5 antagonists or potential HIV-1 entry inhibitors. Furthermore, from the generated atom-based 3D-QSAR model, all of the parameters showed certain reliability and feasible predictability to help us design new and high selectivity CCR5 inhibitors.

The fourth study (Chapter 7), explored the atomistic basis of why the M184I single mutation renders complete resistance of HIV-1 RT to lamivudine. Multiple molecular dynamics simulations, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analyses adequately clarified the effect of the M184I mutation on drug resistance to lamvudine. Results presented in this study verified that M184I mutation decreased drug binding affinity, distorted ligand optimum orientation in RT active site and affected the

overall protein conformational landscape. The results also provided some potential clues for further design of novel inhibitors that are less susceptible to drug resistance.

In the fifth study (Chapter 8), we identified potential HIV-Nef inhibitors by exploiting the structural features of B9 using an integrated computational tools framework. The top identified hit compounds demonstrated comparatively better binding affinities and relatable binding modes compared to the prototype antagonist, B9. Top identified hits were proposed as new potential novel leads targeting HIV-Nef with a detailed analysis of their respective binding modes.

The sixth report (Chapter 9), aimed to reveal the dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations using molecular dynamic simulations. Results verified a more conformational flexible nature of HIV-Nef dimer in the absence of an inhibitor.as compared to B9 bound conformation of HIV-Nef, which was found to be more conformationally rigid with a lesser inter-dimeric association.

We believe that the results obtained from these several studies could be of great benefit in the development of more effective therapeutic interventions for the treatment and cure of HIV/AIDS.

DECLARATION 1 PLAGIARISM

I, **Suri Moonsamy**, declare that:

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

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

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

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written but the general information attributed to them has been referenced
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied 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 (include publications submitted, accepted, in *press* and published).

    Signed ------------------------   **Date** ------------------

DECLARATION 2 – PUBLICATIONS

1. Moonsamy, S. and Soliman, M.E. (2014) Dual Acting HIV Inhibitors: Integrated Rational *in-silico* Design Strategy, *Medicinal Chemistry Research*, 23, 682-689. (***Published***)

   Contribution:

   Moonsamy, S.: contributed to the project by performing all the experimental work and manuscript preparation and writing

   Soliman, M.E.: supervisor

   Appendix A: Pdf version of the publication

2. Moonsamy, S. and Soliman, M.E. (2014) Computer-aided perspective for the design of flexible HIV non-nucleoside reverse transcriptase inhibitors (NNRTIs): de-novo design, virtual screening and molecular dynamics simulations, *Letters in Drug Design and Discovery*, 11, 513-524. (***Published***)

   Contribution:

   Moonsamy, S.: contributed to the project by performing all the experimental work and manuscript preparation and writing

   Soliman, M.E.: supervisor

   Appendix B: Pdf version of the publication

3. Moonsamy, S., Dash, R.C., and Soliman, M.E. (2014) Integrated computational tool for identification of CCR5 antagonists as potential HIV-1 entry inhibitors: homologymodeling, virtual screening, molecular dynamics simulations and 3D QSAR analysis, *Molecules,* 19, 5243-5265. *(**Published**)*

   Contribution:

   Moonsamy, S.: contributed to the project by performing all the experimental work and manuscript preparation and writing

Dash, R.C.: contributed to the project by performing Quantitative Structure Activity Relationship (QSAR) experimental work, data analysis and interpretation of QSAR results

Soliman, M.E.: supervisor

Appendix C: Pdf version of the publication

4. Moonsamy, S., Soumendranath, B., Walker, R.C., and Soliman, M.E. (2014) Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations, *Cell Biochemistry and Biophysics*. ***(Manuscript submitted)***

Contribution:

Moonsamy, S.: contributed to the project by performing experimental work and manuscript preparation and writing, data analysis and interpretation

Soumendranath, B.: contributed to the project by performing experimental work, data analysis and interpretation and writing of results

Walker, R.C.: contributed to the project by providing editing to the Amber sections in the manuscript

Soliman, M.E.: supervisor

5. Moonsamy, S., Soumendranath, B., and Soliman, M.E. (2014) Identification, binding mode and prospective chemical structural features of novel Nef protein inhibitors as potential anti-HIV drugs, *Medicinal Chemistry Research*. ***(Manuscript submitted)***

Contribution:

Moonsamy, S.: contributed to the project by performing experimental work and manuscript preparation and writing, data analysis and interpretation of results

Soumendranath, B.: contributed to the project by performing experimental work, data analysis and interpretation and writing of results

Soliman, M.E.: supervisor

6. Moonsamy, S., Soumendranath, B., and Soliman, M.E. (2014) Dynamic features of apo and bound HIV-Nef protein reveal the anti-HIV dimerization inhibition mechanism, *Journal of Receptors and Signal Transduction. **(Manuscript submitted)***

Contribution:

Moonsamy, S.: contributed to the project by performing experimental work and manuscript preparation and writing, data analysis and interpretation of results

Soumendranath, B.: contributed to the project by performing experimental work, data analysis and interpretation and writing of results

Soliman, M.E.: supervisor

RESEARCH OUTPUT

A. LIST OF PUBLICATIONS

Published

1. Moonsamy, S. and Soliman, M.E. (2014) Dual Acting HIV Inhibitors: Integrated Rational *in-silico* Design Strategy, *Medicinal Chemistry Research*, 23, 682-689.

2. Moonsamy, S. and Soliman, M.E. (2014) Computer-aided perspective for the design of flexible HIV non-nucleoside reverse transcriptase inhibitors (NNRTIs): de-novo design, virtual screening and molecular dynamics simulations, *Letters in Drug Design and Discovery*, 11, 513-524.

3. Moonsamy, S., Dash, R.C., and Soliman, M.E. (2014) Integrated computational tool for identification of CCR5 antagonists as potential HIV-1 entry inhibitors: homology modeling, virtual screening, molecular dynamics simulations and 3D QSAR analysis, *Molecules,* 19, 5243-5265.

B. Submitted

1. Moonsamy, S., Soumendranath, B., Walker, R.C., and Soliman, M.E. (2014) Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations, *Cell Biochemistry and Biophysics.*

2. Moonsamy, S., Soumendranath, B., and Soliman, M.E. (2014) Identification, binding mode and prospective chemical structural features of novel Nef protein inhibitors as potential anti-HIV drugs, *Medicinal Chemistry Research.*

3. Moonsamy, S., Soumendranath, B., and Soliman, M.E. (2014) Dynamic features of apo and bound HIV-Nef protein reveal the anti-HIV dimerization inhibition mechanism, *Journal of Receptors and Signal Transduction.*

C. CONFERENCES

1. Poster presentation "Computer-aided identification of CCR5 antagonists as potential HIV-1 entry inhibitors: homology modeling, virtual screening, molecular dynamic simulations and QSAR analysis" – Molecular Interactions Biomolecules VI Conference, Prague, Czech Republic, September 2013.

2. Oral presentation "An insight into diverse computational techniques for developing effective treatments against HIV" – Three's Company Pharmacy Conference, Cape Town, South Africa, October 2013.

3. Oral presentation "An insight into diverse computational tools for effective therapies against HIV" – CHPC National Meeting 2013, Cape Town, South Africa, December 2013.

ACKNOWLEDGEMENTS

The grace of the Almighty has chaperoned me throughout the course of my studies and my life.

I would like to give thanks to the following individuals:

- Whom have both directly or indirectly supported and assisted me towards the completion of this thesis.

- My supervisor Dr Mahmoud E.S. Soliman for his constant support, guidance invaluable knowledge, and his never-ending motivation, kindness and encouragement. Besides his immerse guidance, he has given me the experience, enhanced my confidence and knowledge tremendously. All of that was very fundamental for a person like myself whom before has never been exposed to computational chemistry. This has enhanced my knowledge of the field of Medical Research, and emerging medical technologies such as molecular modelling and bioinformatics.

- My colleagues, in the UKZN Molecular Modelling Laboratory team (2013-2014) for sharing their knowledge of research methodologies and computational chemistry. Also, their constant support, love and motivation during my PhD.

- The Centre for High Performance Computing (CHPC) for technical support by providing the various clusters and guidance of the helpdesk members.

- The National Research Foundation of South Africa and College of Health Sciences, UKZN for financial support.

- My supervisor and post-doctoral colleagues for their editorial support.

- My parents, Mr Rama and Mrs Roshilla Moonsamy and my twin brother, Strini Moonsamy for their unconditional love, support, motivation and understanding.

- My extended family members and friends for their never-ending love and support during my studies and in all my life's ventures.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3TC	Lamivudine
AIDS	Acquired immune-deficiency syndrome
α	Alpha
Apo	Wild type
Arg	Arginine
Asp	Aspartic acid
β	Beta
C	Carbon
AZT	Zidovudine
CCR5	Chemokine co-receptor 5
DCC	Dynamic-cross correlation
DRV	Darunavir
EFZ	Efavirenz
FDA	USA Food and drug administration
GAFF	General amber force field
Gln	Glutamine
Glu	Glutamic acid
Gp	Glycoprotein
HAART	Highly-active antiretroviral therapy
HBa	Hydrogen bond acceptor
HBd	Hydrogen bond donor
HP	Hydrophobic moieties
His	Histamine
HIV	Human immune-deficiency syndrome
HTS	High-throughput screening
Ile	Isoleucine
IN	Integrase
INIs	Integrase inhibitors

Lys	Lysine
MD	Molecular dynamics
Met	Methionine
MM/GBSA	Molecular mechanics/Generalised-born surface area
Nef	Negative factor protein
NRTIs	Nucleoside reverse transcriptase inhibitors
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
NVP	Nevirapine
PCA	Principle component analysis
PES	Potential energy surface
Phe	Phenylalanine
PLS	Partial least squares
PME	Partial mesh Ewald method
Pro	Proline
PR	Protease
PIs	Protease inhibitors
PDB	Protein data bank
QM/MM	Quantum mechanics/molecular mechanics
QSAR	Quantity structure-activity relationship
RESP	Restrained electrostatic potential
Rg	Radius of gyration
RIN	Residue interaction network analysis
RMSD	Root mean-squared deviation
RMSF	Root mean-squared fluctuations
RPV	Rilpvirine
RT	Reverse transcriptase
RTV	Ritonavir
SASA	Solvent accessible surface area
SBVS	Structure-based virtual screening
Ser	Serine

Thr	Theorine
TMC 278	Rilpvirine
Trp	Tryptophan
Tyr	Tyrosine
UNAIDS	Joint United-nations programme on HIV/AIDS
Val	Valine
VdW	Van der Waal's forces

TABLE OF CONTENTS

CHAPTER 1

1.1. Background and rationale for this study

HIV/AIDS still remains a challenging epidemic that is rapidly characterized by high mortality and morbidity rates. Stringent control of this "global killer" is now perhaps the foremost development imperative for the world. It has been reported that 34 million people live with HIV/AIDS globally [1, 2]. In sub-Saharan Africa, an approximated 22.9 million individuals constitute the overall global estimate [3].

Despite the fact that Highly-Active Antiretroviral Therapy (HAART) has proved successful in order to ensure the efficacy of such a primary therapeutic regimen for AIDS, the benefits of this tactic are more than often compromised by problems, such as patient compliance, complicated dosing, intolerable acute and chronic toxicity profiles that leads to the failure of complete adherence and which in turn results in multi-drug resistance.

In recent years, computational approaches have emerged as close counterparts to experiments in modern drug discovery process and in understanding complex biological phenomena.

Therefore, in one report, a unique strategy was generated for developing dual acting inhibitors against HIV-1 protease (PR) and reverse transcriptase (RT) that may result in lower toxicity profiles, easier dosing regimens and lower or no multi-drug resistance. The unique strategy encompassed "loop-docking" and molecular dynamics simulations.

Although an array of inhibitors haven been approved for targeting RT, their therapeutic effects are hampered by several complications, such as NRTIs exhibit high levels of cellular toxicity [4], compared to NNRTIs, which are severely weakened by the rapid emergence of drug-resistant strains [29, 30]. Therefore, it is suggested that more flexible NNRTIs may be more effective against drug-resistant strains. Bruccoleri [5] proposed the conception of utilizing flexible hydroquinone-based compounds as potential mutant-resistant NNRT inhibitors, however, no experimental or computational evidence supported this proposal. Thus, a study presented herein, provided the computational framework support by querying chemical databases, screening and verifying the

activity of a set of flexible hydroquinone-based compounds as mutant –resistant non-nucleoside reverse transcriptase inhibitors (NNRTIs),

In an effort to combat HIV-1, depends on targeting the critical stages of the viral life cycle, such as viral entry, replication and maturation. Because of the high genetic variability of HIV-1 acquired mutations accumulated during the viral replicative process, this has often resulted in drug resistance. Thus, a better understanding of the mutational effect on the binding conformation is of urgent importance. This may lead to development of antiretroviral drugs that are less prone to resistance.

An important constituent of triple-drug anti-AIDS therapy is the NRTI, lamivudine. It has been proven that single mutations at residue 184 of RT in HIV cause high-level resistance to 3TC and contribute to the failure of anti-AIDS combination therapy. Thus, molecular dynamics (MD) simulations, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analysis were used in a report herein to provide useful information for understanding the drug resistance mechanism against lamivudine. Also, the comprehensive analysis could also provide potential cues for the further design of novel inhibitors that are less susceptible to drug resistance.

Recently acquired knowledge about the HIV entry process points to new strategies to block viral entry. To this end, the homology model of the chemokine co-receptor 5 protein (CCR5) was generated and chemical databases were queried in order to identify novel CCR5 antagonists as potential HIV-1 entry inhibitors.

Because of its major role in HIV-1 pathogenicity, Nef protein has proven to be a very important target in anti-HIV drug design and discovery process. To date, only one small molecule, B9, has been reported as a potent Nef inhibitor. Herewith, exploiting the structural features of B9 and relatable binding modes by using an integrated computational tools framework to identify more potential HIV-Nef inhibitors.

Understanding Nef dimerization process and its inhibition mechanism is crucial for the design of more potent inhibitors as anti-HIV antagonists. Molecular dynamic simulations could be used to reveal the dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations.

The results obtained from these several studies could of great benefit in the development of more effective therapeutic interventions for the treatment and cure of HIV/AIDS.

1.2. Aims and objectives

This study has **six** major aims:

1. To design novel inhibitors with potential dual activity against HIV Protease (PR) and Reverse transcriptase (RT) enzymes. To accomplish this, the following objectives were outlined:

   1.1. To design dual inhibitors using an integrated "tough" *in silico* rational design strategy that exploits the structural features of both the inhibitor and enzyme active site using a "loop docking/MD strategy".

   1.2. To initially design novel leads based on the pharmacophoric features conserved in most potent FDA-approved PR and NRT inhibitors.

   1.3. To subject the pharmacophore-based leads and search the Zinc Database for structure-based scaffolds against the respective enzyme to estimate their binding affinities.

   1.4. To accurately estimate the binding affinity and enhance docking calculations predications by implementing "loop docking".

   1.5. To measure the stability of the resulted docked structures over a reasonable time scale using molecular dynamic simulations. Furthermore, all docked structures that will fail to exhibit stability will be subjected to further docking or rejected.

   1.6. To compute ligand-enzyme interactions on post-dynamic structures to ensure that the proposed leads fit well into the enzyme active site and whether exhibit similar interaction trends as reference drugs.

1.7.   To quantitatively estimate the contribution of each amino acid residue in the enzyme active site towards ligand binding.

1.8.   To screen a wider set of substituents for further lead optimization and expansion.

2. To verify the activity of hydroquinone-based compounds as non-nucleoside reverse transcriptase inhibitors (NNRTIs), as proposed by Bruccoleri [5]. To accomplish this, the following objectives were outlined:

2.1.   To design and investigate the binding affinity and themes for a novel set of different   hydroquinone-based structures against RT enzyme at the molecular level.

2.2.   To identify a wider range of novel hydroquinone-based compounds from the commercially available chemical databases.

2.3.   To perform docking calculations, molecular dynamic stimulations and post-dynamics analysis and per-residue energy contribution in order to accurately estimate the relative binding affinities of the proposed compounds against RT enzyme as well as to ensure the stability of the resulted ligand-enzyme complexes.

2.4.   To validate docking calculations will adopt the same docking approach and apply to different FDA-approved NNRTIs and compare results against experimental data.

3. To identify novel CCR5 antagonists as potential HIV-1 entry inhibitors using integrated computational tools. To accomplish this, the following objectives were outlined:

3.1.   To identify a wider range of constituents with oxamino-piperidino-piperidine amide analogs properties from commercially chemical databases by performing hybrid structure-based and ligand-based virtual screening.

3.2.   To validate docking calculations will adopt the same docking approach and apply to different FDA-approved NNRTIs and compare results against experimental data.

3.3.   To construct a homology model of CCR5 enzyme using CXCR4 has a structural template and CCR5 human protein sequence.

3.4.   To accurately estimate the binding affinities using docking calculations and estimate the stability of systems over a specified time scale using molecular dynamic simulations. As well as to investigate the nature of the overall interaction themes between the generated leads and target protein and specific amino acid involved in ligand binding using post-dynamic analyses.

3.5.   To build and generate an atom-based 3D-QSAR model using a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC_{50} (mM) data taken from literature.

4. To provide a more comprehensive insight into the precise impact of the M184I mutation on RT resistance to lamivudine. To accomplish this, the following objectives were outlined:

4.1.   To provide an deeper understanding into the drug resistance mechanism of the RT-M184I mutant to lamivudine using validated models of wild type and M184I mutant lamivudine–RT complexes; using multiple molecular dynamic simulations, post-dynamic analysis, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analysis, respectively.

4.2.   To construct validated the 3D X-ray crystal structures of lamivudine-RT complexes using a "ligand fitting" strategy.

5. To identify more potent potential HIV-Nef inhibitors by exploiting the structural features of B9 using an integrated computational tools framework. To accomplish this, the following objectives were outlined:

5.1.   To identify a wider range of constituents with diphenylpyrazole analog properties to B9 from commercial chemical databases by performing hybrid shape similarity and pharmacophore-based virtual screening.

5.2.   To estimate different binding energy affinities and poses/trends using docking calculations.

5.3.   To validate docking poses and binding energy trends using molecular dynamic simulations as well as to provide further verification on the resulted- docked systems.

5.4.   To quantify individual amino acid interactions towards total binding free energy computed per-residue interactions using MM/GBSA approach. Also, to provide a set of chemical structures and detailed binding mode analysis for further optimization and expansion.

6. To reveal the dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations. To accomplish this, the following objectives were outlined:

6.1.   To understand the dynamic landscape of HIV-Nef protein structure, dynamics, dimerization and inhibitory mechanisms using classical comparative molecular dynamic simulations for the apo and bound protein. This was also assisted by a wide range of post-dynamic analyses, which includes: dynamic cross correlation (DCC), principal component analysis (PCA), radius of gyration (Rg), protein mobility plots as well as monitoring other several metrics.

1.3.   Novelty and significance of this study

This incapacitating disease produces high viral titers and in turn causes destruction to the host immune system and eventually results in death of infected individuals. The HIV viral core encloses three crucial enzymes, reverse transcriptase (RT), integrase (IN) and protease (PR), which are imperative targets in viral replication and survival [6-9]. Reverse transcriptase (RT) is a vital component in the HIV-1 life cycle, where this multifunctional enzyme facilitates the conversion of single-stranded viral RNA genome into double-stranded DNA (dsDNA)[10, 11]. HIV-1 IN catalyzes two step-wise reactions that are imperative in viral integration: catalysis of virus-host dsDNA, endonucleolytic processing of the 3' ends of HIV-1 dsDNA and strand transfer of viral and cellular DNA [12, 13]. However, HIV-1 PR is responsible for catalytic cleavage of newly synthesized polypeptide precursor proteins, *gag* and *gag-pol*, which are essential elements for virion maturation [14, 15].

The HIV-1 life cycle presents a myriad of factors and enzymes to be targeted for therapeutic intervention. Despite the ongoing HIV/AIDS research, to date no cure yet exists for this disease. Currently, the most effective treatment comprises of a complex "cocktail" of Food and Drug Administration (FDA)-approved and clinical trial drugs that involve multiple drug targets [16-20]. These drugs target and attack the virus at different stages of its lifecycle, thereby halting viral replication and reducing destruction of the immune system. They include: protease inhibitors (PIs), reverse transcriptase (RT) inhibitors, integrase (IN) inhibitors and entry inhibitors [21, 22].

As stated earlier, three main HIV-1 enzymes, protease (PR), integrase (IN) and reverse transcriptase (RT), have been exploited as the core of chemotherapy for AIDS treatment [16-20]. Although, researchers have designed inhibitors that successfully target and inhibit each enzyme, however, the therapeutic effects of these drugs are severely compromised by numerous implications. For instance, NRTIs have been reported to demonstrate high intrinsic toxicity [4], whilst, the less toxic PIs, INIs and NNRTIs are drastically hampered owing to the rapid emergence of drug-resistant viral strains [23, 24]. Furthermore, persistent therapeutic profiles can also result in chronic toxicity.

At present, due to extensive research, it is known that a single therapy might not always provide the desired efficacy vs side effect paradigm. When the high specificity proves insufficient for a single target; medicinal chemists and other pharmaceutics have resorted to molecular modelling drug design techniques. Also, presently a common tactic employed by clinicians to overcome HIV disease resistance to a single drug is the use of multidrug therapy. HAART combines two or more antiretroviral drugs that inhibits two HIV targets, thereby blocking viral fusion and which in turn effectively reduces viral load and helps improves a patient's life [25-28]. Despite the fact that HAART has proved successful in order to ensure the efficacy of such a primary therapeutic regimen for AIDS, the benefits of this tactic are more than often compromised by problems, such as patient compliance, complicated dosing, intolerable acute and chronic toxicity profiles that leads to the failure of complete adherence and which in turn results in multidrug resistance.

The concept of designing multifunctional inhibitors has been proposed before [29-32], where medicinal chemists and other researchers highlight the increasing shift towards multifunctional inhibitors as well as the various opportunities and advantages allied with the development of such compounds.

Herein, we reported the main concept of designing potential proposed dual acting PR/RT HIV inhibitors assisted by an integrated *in silico* computational approach that exploited the structural features of both the inhibitor and enzyme active site. We introduced an *in silico* "loop-docking/MD strategy" that relied on the use of docking calculations to predict the preferred binding orientation of potential drug candidates within the enzyme active sites, then followed by molecular dynamic simulations to accurately estimate the binding affinity, and the stability of inhibitor-enzyme complexes and per-residue interaction binding analysis.

Although, a large number of inhibitors that target RT have been approved for treatment [22, 33], there are many complications that hinder the therapeutic effect of these inhibitors. For instance, NRTIs demonstrated high levels of cellular toxicity [4], whereas the highly potent and less toxic NNRTIs are severely weakened by the rapid emergence of drug-resistant strains [23, 24]. Therefore, tackling the drug-resistant implications associated with NNRTIs could result in less toxic and more potent inhibitors as anti-HIV RT. It is strongly believed that more flexible NNRTIs could be more effective against drug-resistant mutants.

The concept of designing flexible RT inhibitors has been proposed before [34, 35]. Ohtaka and his colleagues made a conclusion that a protease inhibitor required to be flexible in nature in order to bind to mutant enzymes [34]. Soon afterwards, Das *et al* (2005), concluded that conformational flexibility of an NNRTI was crucial in preserving the functionality of an NNRTI being exposed to mutations of resist antiviral strains [35].

The proposal of hydroquinone-based structures as flexible NNRTIs was first introduced by Bruccoleri [5], however, this proposal was not supported by either experimental or computational evidences. Therefore, the work reported herein is an attempt to verify the activity of hydroquinone-based compounds as NNRTIs, as proposed by Bruccoleri [5]. In order to accomplish this, we have designed and investigated the binding affinity and interaction themes for a novel set of different hydroquinone-based structures against the RT enzyme at the molecular level.

An important fact to consider when predicting the efficacy of a library of ligands is the time needed to calculate the affinity of a typical ligand. Screening of large virtual libraries demands a high throughput of ligands and requirement for a short period of time in evaluating a single compound. Varieties of computational approaches have been developed over the years to predict binding constants and this from statistical ones to methodologies based on the evaluation of

binding energies and interaction networks. Structure-based virtual screening was performed to identify a wider range of novel hydroquinone-based compounds from the commercially available chemical databases then followed this de-novo design. Furthermore, we performed docking calculations, molecular dynamic stimulations, post-dynamics analysis and per-residue energy contribution analysis.

Even though NRTIs have served as the cornerstones of successful HIV therapy, the largest problem in HIV drug therapy is that the virus mutates very quickly, leading not only to drug resistance [36], but also drug-resistant variants of the virus that have mutations in the RT target protein [37, 38]. For this reason a molecular understanding of the impact of mutations on drug resistance will assist in the design of more potent drugs that are active against resistant strains.

An important constituent of triple-drug anti-AIDS therapy is the NRTI 2', 3'-dideoxy-3'-thiacytidine (3TC, lamivudine). It has been proven that single mutations at residue 184 of RT in HIV cause high-level resistance to 3TC and contribute to the failure of anti-AIDS combination therapy [39]. Partially due to the lack of an X-ray crystal structure of the lamivudine-RT complex, the precise mechanism by which the M184I mutant develops resistance to lamivudine is not fully understood. Based on a previously built model from "unbound" X-ray crystal structures of the wild type and M184I mutant, it is thought that steric conflict between the oxathiolane ring of lamivudine and the side chain of beta-branched amino acids Ile at position 184 perturbs inhibitor binding, leading to a reduction in incorporation of the analog [39]. However, this assumption was based on approximated "fitted" and "static" models.

Herewith, we believe this to be a first account study, where multiple molecular dynamics simulations, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analysis were employed in order to investigate the atomistic basis of the mystery of why the M184I single mutation leaves the HIV-1 reverse transcriptase (RT) totally resistant to lamivudine. We believe that, to develop potent and effective anti-HIV NRTIs against the viral variants, understanding of the molecular basis of the M184I RT mutation at atomic level will be very critical.

Recently acquired knowledge about the HIV entry process points to new strategies to block viral entry. For most HIV strains, the successful infection of their target cells is mainly dependent on the presence of CD4 surface molecule, which serves as the primary virus receptor [40, 41]. One such

example is the chemokine receptor CCR5 has proven to be a crucial pharmaceutical target in the contexts of HIV-1 and other inflammatory diseases [42-45].

Only one small molecule, maraviroc, has been currently approved as a potent CCR5 antagonist by the FDA [46-48]. Despite the fact that no structural information exists about CCR5 precise binding site and no reported X-ray 3D crystal structure of CCR5 is yet available; only three other previous studies have reported the concept of designing small molecule CCR5 antagonists [42-45]. For the purpose of our study, we created and validated an actual CCR5 homology model using CXCR4 crystal structure and maraviroc as structural templates. For the work described herein, we attempted to report more potential small molecule CCR5 antagonists via a validated CCR5 crystal structure together with an integrated computational design of a hybrid structure-based and ligand-based virtual screening, extensive molecular dynamic simulations and post-dynamic analyses.

We further also included another highly prominent statistical computational approach to predict binding efficacy of constants using QSAR. This method attempts to correlate molecular properties to physiochemical properties of a set of structures to predict biological activities of other compounds, thereby reducing the time spent on synthesizing numerous compounds. Therefore, we incorporated a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC50 (mM) data taken from literature for the development of an atom-based 3D-QSAR model.

As an emerging target, HIV-Negative Factor (Nef) protein, an accessory pathogenic factor, plays a significant role in HIV replication [49, 50]. Reports have indicated that removal of Nef protein leads to high viral insensitivity levels. Therefore, targeting Nef is considered a key strategy towards HIV/AIDS treatment.

Up-to-date, only one compound has been reported as a potent Nef antagonist, known as B9 [109, 110]. This has prompted us to provide this first account of an integrated computational framework in order to identify more potential Nef inhibitors. Herein, using a hybrid ligand virtual screening, shape similarity- and pharmacophore-based, approach combined with molecular dynamics and

post- dynamics analysis were applied to identify potential new leads targeting HIV-Nef with a detailed analysis of their binding modes.

Because of its major role in HIV-1 pathogenicity, Nef protein has been proved to be a very important target in anti-HIV drug design and discovery process [2, 50]. Understanding Nef dimerization process and its inhibition mechanism is crucial for the design of more potent inhibitors as anti-HIV antagonists. In recent years, molecular dynamics simulations and post dynamics calculations emerged as a close counterpart to experiment and helps in understanding complex biological phenomena.

Herein, in order to understand the dynamic landscape HIV-Nef dimerization process and its inhibition, comparative molecular dynamics simulations were employed for the apo and bound protein. A wide range of post-dynamic analyses were carried out in order to accomplish this task – these include; dynamic cross correlation (DCC), principal component analysis (PCA), radius of gyration (Rg), protein mobility plots as well as monitoring other several metrics. To our best knowledge, this is the first account of such comprehensive computational study on this crucial HIV target. Therefore, we believe that this report serves as a cornerstone towards the understanding HIV-Nef protein structure and dynamics and its inhibitory mechanism.

To this end, the work presented in this thesis is considered a key cornerstone towards further understanding of HIV therapy, in general, and the South African version, in particular. Also, this study could serve as a road map for further HIV drug design and development.

1.4. Overview of this work

This thesis is divided into **ten chapters**, including this one:

Chapter 1: This chapter addresses the background, aim and objectives, significance as well as the general outline and structure of the thesis.

Chapter 2: It provides a general overview on the HIV/AIDS epidemic and therapy. The chapter begins with a historic background on HIV/AIDS epidemic followed by some updated statistics on the number of HIV infected individuals worldwide and in Africa, respectively. The chapter also highlights numerous aspects such as HIV virus structure, life cycle, and the important enzymes required for virus maturation, HIV drug targets, drug resistances, drug inhibitors, different mechanisms of action and side effects. The HIV PR, RT and Nef enzymes, which are the crucial drug targets and the main focus of the work, are then addressed in details including its structures and functions, diverse mechanisms of actions, molecular dynamics of these drug targets, inhibitor design strategies and the currently approved FDA drugs.

Chapter 3: This chapter provides a brief introduction to computational chemistry, diverse molecular modeling and molecular simulations techniques and their respective applications. Some computational approaches have been theoretically explained. This is followed by a highlight on the different computational tools used in HIV research with main focus on molecular dynamics simulations, molecular docking, principle component analysis, residue interaction network analysis, hybrid quantum mechanics and molecular mechanics (QM/MM), binding free energy calculations and subsequent dynamics calculations.

Chapter 4: (Published work)

This chapter demonstrates a unique strategy for developing dual acting inhibitors against HIV-1 protease and reverse transcriptase via an integrated "tough" *in-silico* computational protocol that exploits the structural features of both the inhibitor and enzyme active sites. The computational technique implemented in this study relies on "loop docking" and molecular dynamic simulations.

Chapter 5: (Published work)

The proposal of hydroquinone-based structures as flexible non-nucleoside reverse transcriptase inhibitors (NNRTIs) was first introduced by Bruccoleri [5], however, this proposal was not supported by either experimental or computational evidences. Therefore, in this chapter, the work reported herein is an attempt to verify the activity of hydroquinone-based compounds as NNRTIs, as proposed by Bruccoleri [5]. In order to accomplish this, we have designed and

investigated the binding affinity and themes for a novel set of different hydroquinone-based structures against RT enzyme at molecular level.

Chapter 6: (Published work)

This chapter highlights using integrated computational tools including via a hybrid structure-based and ligand-based virtual screening approach, to identify novel CCR5 antagonists as potential HIV-1 entry inhibitors. Since the X-ray 3D CCR5 crystal structure is not yet available, the actual homology model of CCR5 was created using CXCR4 crystal structure as a structural template. The structure and dynamics of CCR5 as well as the nature of the overall binding affinity themes between the generated leads and target protein were studied using molecular dynamics followed by extensive post-dynamic analysis. Also included in this chapter, a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC$_{50}$ (mM) data taken from literature for the development of an atom-based 3D-QSAR model.

Chapter 7: (Manuscript submitted)

This chapter aims to provide more comprehensive insight into revealing the mystery of the M184I single mutation that leaves the HIV-1 reverse transcriptase (RT) totally resistant to lamivudine. The computational techniques included multiple molecular dynamics simulations, binding free energy calculations, principle component analysis (PCA), and residue interaction network (RIN) analysis. We believe that, to develop potent and effective anti-HIV NRTIs against the viral variants, understanding of molecular basis of M184I RT mutation at atomic level will be very critical.

Chapter 8: (Manuscript submitted)

This first account report of an integrated computational framework used in order to identify more potential HIV-Nef inhibitors by exploitation of the only reported potent Nef inhibitor, B9 structural features. The computational tools include: hybrid shape similarity-and-pharmacophore-based virtual screening approach, molecular dynamics simulations, post-dynamics analysis and binding free energy calculations. Also, a set of chemical structural features and a detailed analysis of the new leads respective binding modes is provided for further lead optimization and expansion.

Chapter 9: (Manuscript submitted)

This is the first effort of such a comprehensive computational reported study on the crucial HIV target Nef in order to understand the dynamic landscape of HIV-Nef dimerization process and its inhibition using comparative molecular dynamic simulations and a variety of post-dynamic analyses (i.e. dynamic cross correlation (DCC), principal component analysis (PCA), radius of gyration (Rg), protein mobility plots as well as monitoring other several metrics) for both apo and inhibitor bound protein states. We believe that this report serves as a cornerstone towards the understanding HIV-Nef protein structure and dynamics and its inhibitory mechanism.

Chapter 10: This chapter expounds the overall concluding remarks of the entire thesis and future plans and recommendations.

1.5. References

1. Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Medicinal Chemistry Research*.

2. Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. . (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants, *Retrovirology* 1-23.

3. 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 & Drug Design 81*, 208-218.

4. Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. . (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses, *Antiviral Research 71*, 256.

5. Bruccoleri, A. (2013) Positional Adaptability in the Design of Mutation-Resistant Nonnucleoside HIV-1 Reverse Transcriptase Inhibitors: A Supramolecular Perspective., *AIDS Research and Human Retroviruses 29*, 1-9.

6. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV, *Journal of Molecular Biology 285*, 1-32.

7. Levy, J. A. (2007) *HIV and the Pathogenesis of AIDS, 3rd Edition*, Amer Soc Microbiology, 1752 N Street Nw, Washington, Dc 20036-2904 USA.

8. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein, *Cell 89*, 263-273.

9. McKeage, K., and Perry, C. M. (2002) Trastuzumab - A review of its use in the treatment of metastatic breast cancer overexpressing HER2, *Drugs 62*, 209-243.

10. Hu, W.-S., and Hughes, S. H. (2012) HIV-1 reverse transcription, *Cold Spring Harbor perspectives in medicine 2*.

11. Craigie, R., and Bushman, F. D. (2012) HIV DNA Integration, *Cold Spring Harbor perspectives in medicine 2*, a006890.

12. Zheng, Y. H., Lovsin, N., and Peterlin, B. M. A. (2005) Newly identified host factors modulate HIV replication, *Immunol. Lett. 97*, 225-234.

13. Karn, J., and Stoltzfus, C. M. (2012) Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression, *Cold Spring Harbor perspectives in medicine 2*, a006916.

14. Spearman, P. (2006) Cellular cofactors involved in HIV assembly and budding, *Current Opinion in HIV and AIDS 1*, 200-207.

15. Brik, A., and Wong, C. H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & Biomolecular Chemistry 1*, 5-14.

16. Johnson, B. C., Metifiot, M., Ferris, A., Pommier, Y., and Hughes, S. H. (2013) A Homology Model of HIV-1 Integrase and Analysis of Mutations Designed to Test the Model, *Journal of molecular biology 425*, 2133-2146.

17. Zhan, P., Chen, X., Li, D., Fang, Z., De Clercq, E., and Liu, X. (2013) HIV-1 NNRTIs: structural diversity, pharmacophore similarity, and impliations for drug design, *Medicinal Research Reviews 33*, E1-E72.

18. Patel, J. R., and Prajapati, L. M. (2013) Predictive QSAR modeling on tetrahydropyrimidine-2-one derivatives as HIV-1 protease enzyme inhibitors, *Medicinal Chemistry Research 22*, 2795-2801.

19.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets: Infectious Disorders 2*, 17-32.

20.   Morah, E. U. (2007) Are People Aware of Their HIV-positive Status Responsible for Driving the Epidemic in SubSaharan Africa? The Case of Malawi In *Development Policy Review*, pp 215-242.

21.   Pani A, L., A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. . (2002) Targeting HIV: Old and New Players, *Current Drug Target -Infectious Disorders 2*, 17-32.

22.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets - Infectious Disorders 2*, 17-32.

23.   Wainberg MA, S. J., Montaner JS, Murphy RL, Kuritzkes DR, and Raffi F. (2005) Challenges for the clinical development of new nucleoside reverse transcriptase inhibitors for HIV infection., *Antiviral Therapy 10*, 13-28.

24.   Gallant, J. E., Gerondelis, P.Z., Wainberg, M.A., Shulman, N.S., Haubrich, R.H., St Clair, M., Lanier, E.R., Hellmann, N.S., and Richman, D.D. (2003) Nucleoside and nucleotide analogue reverse transcriptase inhibitors: A clinical review of antiretroviral resistance., *Antiviral Therapy 8*, 489–506.

25.   Shehu-Xhilaga, M., Tachedjian, G., Crowe, S. M., and K, K. (2005) Antiretroviral compounds: mechanisms underlying failure of HAART to eradicate, *D - 9440157 12*, 1705-1719.

26.   Cohen, J. (2002) Therapies. Confronting the limits of success, In *Science.*, pp 2320-2324.

27.   Bodiwala, H. S., Sabde, S., Gupta, P., Mukherjee, R., Kumar, R., Garg, P., Bhutani, K. K., Mitra, D., and Singh, I. P. (2011) Design and synthesis of caffeoyl-anilides as portmanteau inhibitors of HIV-1, *D - 9413298 19*, 1256-1263.

28.   Christophe Marchand, J. A. B., Antony Wamiru, Scott Budihas, Ute Möˆlmann, Lothar Heinisch, John W. Mellors, Stuart F. Le Grice and Yves Pommier. (2007) Madurahydroxylactone Derivatives as Dual Inhibitors of Human Immunodeficiency Virus Type 1 Integrase and RNase H, *Antimicrobial agents and chemotherapy 52*, 361-364.

29. Wang, Z., and Vince, R. (2008) Synthesis of pyrimidine and quinolone conjugates as a scaffold for dual *D - 9107377 18*, 1464-3405.

30. Wang, Z., and Vince, R. (2008) Design and synthesis of dual inhibitors of HIV reverse transcriptase and integrase, *D - 9413298 16*, 1464-3391.

31. Espinoza-Fonseca, L. M. (2006) *14*, 897.

32. Blake, L., and Soliman, M. E. S. (2013) Bifunctional Anti-HIV/TB Inhibitors: Perspective from In-Silico Design and Molecular Dynamics Simulations, *Bentham 10*, 712-716.

33. Bauman, J. D., Das, K., Ho1, W.C., Baweja, M., Himmel, D.M., Clark Jr,A.D., Oren, D.A., Boyer, P.L., Hughes, S.H., Shatkin, A,J., and Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design., *Nucleic Acids Research 36*, 5083–5092.

34. Ohtaka H, V.-C. A., Xie D, and Freire E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease., *Protein Sci. 11*, 1908–1916.

35. Das K, L. P., Hughes SH, and Arnold E. (2005) Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors., *Prog Biophys Mol Biol 88*, 209–231.

36. Hamers, R. L., Kityo, C., Sigaloff, K. C., and de Wit, T. F. R. (2013) Pretreatment HIV-1 drug resistance in Africa, *The Lancet infectious diseases 13*, 476-476.

37. Wainberg, M. A., and Turner, D. (2004) Resistance issues with new nucleoside/nucleotide backbone options, *Jaids-Journal of Acquired Immune Deficiency Syndromes 37*, S36-S43.

38. Gao, H. Q., Boyer, P. L., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2000) The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase, *Journal of Molecular Biology 300*, 403-418.

39. Sarafianos, S. G., Das, K., Clark, A. D., Ding, J. P., Boyer, P. L., Hughes, S. H., and Arnold, E. (1999) Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids, *Proceedings of the National Academy of Sciences of the United States of America 96*, 10027-10032.

40. Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major, *Science. 270*, 1811-1815.

41. Fano, A., Ritchie, D. W., and Carrieri, A. (2006) Modeling the structural basis of human CCR5 chemokine receptor function: From homology model building and molecular dynamics validation to agonist and antagonist docking, *Journal of Chemical Information and Modeling 46*, 1223-1235.

42. Perez-Nueno, V. I., Ritchie, D. W., Rabal, O., Pascual, R., Borrell, J. I., and Teixido, J. (2008) Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking, *Journal of Chemical Information and Modeling 48*, 509-533.

43. Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., and Igglessi-Markopoulou, O. (2006) Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques, *Journal of Computer-Aided Molecular Design 20*, 83-95.

44. Aher, Y. D., Agrawal, A., Bharatam, P. V., and Garg, P. (2007) 3D-QSAR studies of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas as CCR5 receptor antagonists, *Journal of Molecular Modeling 13*, 519-529.

45. Kellenberger, E., Springael, J.-Y., Parmentier, M., Hachet-Haas, M., Galzi, J.-L., and Rognan, D. (2007) Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening, *Journal of Medicinal Chemistry 50*, 1294-1303.

46. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Research 39*, D1035-D1041.

47. Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Research 36*, D901-D906.

48. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Research 34*, D668-D672.

49. Sarafianos, S. G., Das, K., Hughes, S. H., and Arnold, E. (2004) Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases, *Current Opinion in Structural Biology 14*, 716-730.

50. Das, S. R., and Jameel, S. (2005) Biology of the HIV Nef protein, *D - 0374701 121*, 315-332.

CHAPTER 2

2. Background on HIV/AIDS

2.1. Introduction

This chapter briefly explains the background, life cycle and different past and present therapeutic regimens for Human Immunodeficiency Virus (HIV), the causative agent of Acquired Immunodeficiency Syndrome (AIDS).

2.2. The current status of global HIV/AIDS

Ever since the first case of illness was reported in 1981, the origin of the Human Immunodeficiency Syndrome (HIV), the causative agent of Acquired Immunodeficiency Syndrome (AIDS) has been the subject of intense challenges and discussions [1]. Since the discovery of HIV, an increased amount of efforts has been directed towards providing a deeper understanding into the cause of this disease, learning more about its invasive and replicative behaviours in host cells.

Since the beginning of this global endemic, over 60 million people have already contracted HIV, whereas, an estimated 30 million individuals died because of HIV-related causes [2]. According to recent statistics, this "global killer" has already infected more than 40 million individuals and proclaimed the lives of an estimated 22 million globally [3]. During 2012, approximately 9.7 million individuals were reported to have received antiretroviral therapy (ART) in both developed and developing countries [4]. The 2013 UNAIDS report [5], reported that an estimated 5.7 million people were living with HIV/AIDS in South Africa. Of this global estimate, 3.3 million infected individuals are under the age of 15, whereas approximately 2.5 million newly infected cases and 1.7 million reported deaths [6, 7]. Currently, an estimated 34 million are living with HIV/AIDS worldwide [6, 7], with an approximated 22.9 million in sub-Saharan Africa that constitute this overall global estimate [8].

To date, we still have no precious findings about whom was first infected with HIV and exactly how this disease was contracted or became widespread from the initial individual. Nevertheless,

conceivably a more demanding concern for researchers today should not be how the AIDS epidemic was born, but rather how those it affects can be treated, how the further spread of HIV can be prevented and how more effective and safer drugs can be designed and developed to beat this vicious virus.

2.3. The quest for a better understanding of the HIV virus

2.3.1. HIV virus: structure and life cycle

Amongst the classified infectious diseases, the Human Immunodeficiency Virus (HIV) infection, the causative agent of AIDS, remains a challenging epidemic [9]. This incapacitating disease produces high viral titers and in turn causes destruction to the host immune system and eventually results in death of infected individuals.

To date, since the discovery of HIV and its association to AIDS, remarkable efforts have been made in order to provide deeper insight into the major physiological and molecular mechanisms underlying this infectious disease. A deeper understanding of the HIV structure and life cycle is fundamental to scientists in order to design potential retroviral inhibitors as well as explore the structural implications of the HIV drug resistance mutations. The HIV virus is highly mutable - different variants of the HIV virus can evolve within the body of a single infected individual [10].

Virus strains are classified into different types, groups and subtypes depending on their genetic similarities [11, 12]. The major issue for effective treatment is attributed to the high genetic diversity amongst HIV variants. HIV is classified as a member of the *Lentivirus* subfamily of retroviruses. HIV is further divided into two subtypes, namely HIV-1 and HIV-2. Although both viruses are responsible for causing AIDS, HIV-1 is primarily responsible for the global HIV epidemic, whereas HIV-2 is located more geographically and has a lower transmission rate and is less virulent as compared to HIV-1 and does not necessarily cause AIDS in all infected persons [13, 14]. Furthermore, HIV-1 can be divided into M (main), O (outlier) and N (non-M or non-O) groups. The M group is commonly further subdivided into subtypes (A, B, C, D, F, G, H, J and K), unique and circulating recombinant forms [13-16]. The genetic material of most organisms are stored in long DNA strands, however, in the case of retroviruses the genetic material is encoded in the

viral Ribonucleic Acid (RNA).Essentially, a retrovirus uses the reverse transcriptase enzyme to reverse transcribe its own RNA into viral DNA within the host cell. Thereafter, the integrase enzyme incorporates the newly generated DNA is into the host genome. By now, the host cell perceives the viral DNA as its own and hence uses it to produce new viral proteins using the protease enzyme that is required for viral replication and infection [17].

2.3.2. Structure of the HIV virus

The HIV virus genome contains two chains of positive single-stranded RNA, which is imperative in the replicative process and to sustain persistent infection (**Figure 2.1.**). The basic structure of HIV consists of three parts: the viral envelope, HIV matrix and the viral core (**Figure 2.1.**). The viral envelope is made up of a lipid bilayer that is derived from the host cell during the bussing process and also contains other proteins embedded in it which forms protrusions. These proteins include: the glycoprotein (gp) 120 required for host cell attachment and the transmembrane gp41 is essential for cell fusion [18-20]. The HIV matrix consists of different proteins and is positioned between the bilayer lipid membrane and the viral core, and also contains the p17 protein, which is an important structural protein involved in all stages of the virus life cycle and provides more stability to virion particle (**Figure 2.1.**) [18-20]. The viral core contains the viral capsule protein p24 (encloses the two ssRNA) and all three vital proteins (reverse transcriptase, protease and integrase) required for viral replication (**Figure 2.1.**). Likewise, the viral core also contains the three main viral genes (*gag, pol and env*), which holds the genetic information that is needed in the production of new structural viral protein particles.
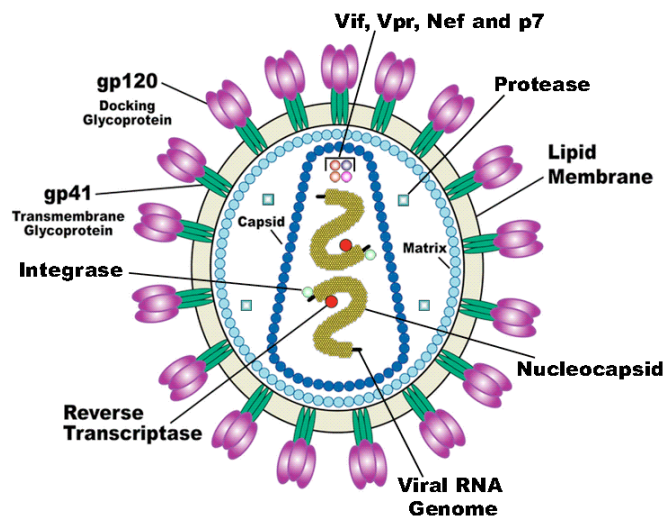
Figure 2.1. Structure of the HIV-1 virus adapted from [21].

2.3.3. HIV life cycle

HIV-1 viral replication is described as a series of steps which begins from host cell attachment, then viral genome integration and ending with the generation of nascent viral particles that eventually mature and then buds off as infectious virions from the newly infected host cell. To date, the armamentarium of antiviral drugs has included a new class of therapeutic regimen drugs for HIV-1 called antiretroviral drugs. As explained in the following text, the HIV-1 life cycle presents a myriad of factors and targets to be targeted for therapeutic intervention.

A typical HIV life cycle is depicted in **Figure 2.2**. HIV-1 primarily infects $CD4^+$ T lymphocytes, dendritic cells and macrophages found in the immune system [22, 23]. In order to have persistently effective infection, HIV-1 requires attachment with viral and host cell membranes (**Figure 2.2.**) [24, 25]. So HIV-1 enters a $CD4^+$ T-cell or a macrophage when the viral surface HIV-1 glycoprotein gp120 binds to a chemokine receptor, either CXCR4 or CCR5 on the host cell surface, thus allowing for viral fusion and entrance into the host cell [19, 26-29]. The series of receptor binding stages triggers conformational changes in the viral envelope, allowing for gp41 hydrophobic domain exposure that mediates fusion with the cellular membrane (**Figure 2.2.**).
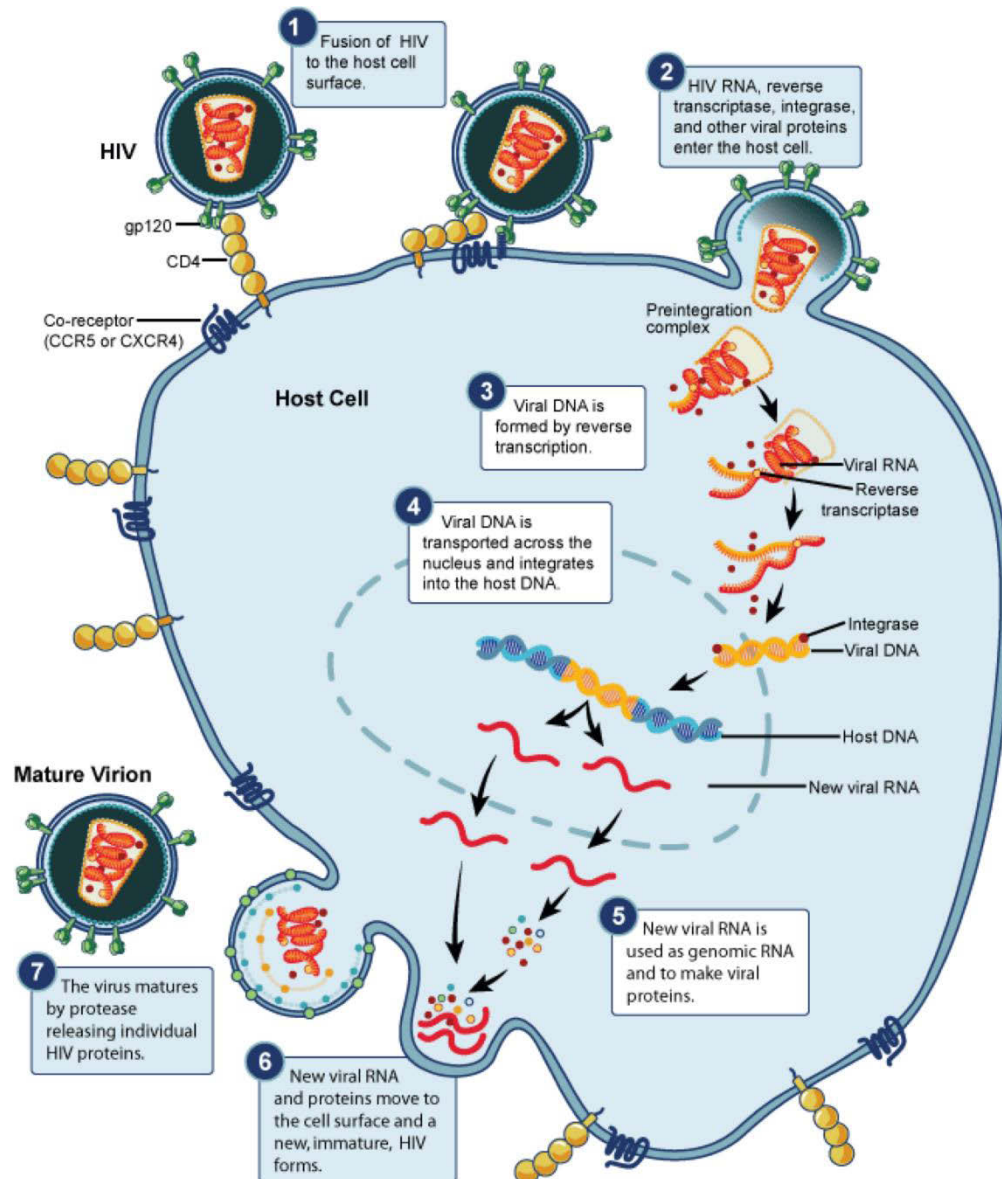
Figure 2.2. Schematic representation of the HIV-1 life cycle [22].

Once the virus successfully binds to the host cell membrane, thus the fusion of the viral envelope with the host cell membrane permits the viral core to become uncoated and all of the viral core components (**Figure 2.2.**), such as viral RNA and different enzymes (reverse transcriptase, protease and integrase) are injected into the cell.

Reverse transcriptase (RT) is a multifunctional enzyme that has RNA-dependent DNA polymerase, RNase-H and DNA-dependent DNA polymerase activities, all of which are crucial for viral replication. RT is responsible for converting ssRNA of HIV-1 into double-stranded DNA

(dsDNA) in the host cytoplasm by a process called reverse transcription (**Figure 2.2.**) [17]. Reverse transcription is known to produce irreversible errors, thus producing frequent mutations within each new virus generation. Also, the end product of reverse transcription is the formation of the viral pre-integration complex (PIC). This pre-integration complex contains both viral and cellular constituents and is translocated to the nucleus, where integrase incorporates the viral dsDNA into the host DNA in a process called integration and this is a step crucial in HIV replication [30].

HIV-1 integrase catalyzes two step-wise reactions that are fundamental in viral integration: catalysis of the virus-host dsDNA, endonucleolytic processing of the 3' ends of HIV-1 dsDNA and strand transfer of viral and cellular DNA [31]. Integration is a point of no return for a cell that fundamentally becomes the "master carrier" of the viral genome, and also known as the provirus (**Figure 2.2.**) [32]. Furthermore, integration can be regarded in part responsible for persistent viral infections.

After successful integration of the viral DNA, the host cell now becomes latently infected with the virus and this provirus anticipates activation. Once the host cell (i.e. the provirus) becomes activated, transported to the nucleus, viral DNA is transcribed into messenger RNA (mRNA) catalyzed by DNA polymerase in a process called transcription (**Figure 2.2.**). The mRNA, a single-stranded DNA strand complementary to the 3' end of the viral DNA strand, carries genetic material from the nucleus to the host cytoplasm and is translated into viral proteins. The end products of translation, cellular enzyme (*env*) precursor polyproteins and transmembrane protein (*gag and pol)* precursors, are then translocated to the cell membrane and assemble into a bud conformation.

When all of the different viral subunits have been generated and processed, they have to be cleaved/separated in order for final assembly into a new virus (**Figure 2.2.**) [33]. This separation or cleavage is controlled by HIV protease enzyme, which cleaves the newly formed polypeptides, and allows for binding to other new functional viral RNA proteins. The final step in HIV-1 life cycle is the budding process (**Figure 2.2.**) [34]. In this process, the genetic material (i.e. the viral RNA and proteins) enclosed in a nucleocapsid fuses with the deformed cell membrane developing a new viral outer membrane [35]. Thereafter, the newly formed HIV virions buds off

from the host cell membrane and is released into bloodstream ready to scavenge and infect other potential host cells (**Figure 2.2.**) [36-38].

The HIV-1 life cycle has an imperative role in viral reproduction and uptake. Understanding the structural dynamics as well as the mechanistic behavior of the HIV-1 life cycle is critical in the developmental of anti-HIV drugs. Most of these drugs impedes the early steps of the HIV-1 life cycle (i.e. attachment and fusion stages) as well as target and inhibit the vital target proteins, reverse transcriptase, integrase and protease and others, critical for viral replication, maturation and survival. These HIV targets will further be discussed in the succeeding sections.

2.4. HIV enzymes as potential drug targets

The viral core encloses three crucial enzymes, reverse transcriptase (RT), integrase (IN) and protease (PR), which are fundamental in viral replication and survival. However, other factors imperative for viral replication such as entry glycoprotein chemokine receptors (e.g. CCR5) and transcription accessory proteins (e.g. Nef) have become targets in drug development in the fight against HIV/AIDS. Therefore, in this section the structural as well as the mechanistic features of these HIV targets will be explained.

2.4.1. Reverse transcriptase (RT)

Reverse transcriptase (RT) is a vital component in the HIV-1 life cycle, where this multifunctional enzyme facilitates the conversion of single-stranded viral RNA genome into double-stranded DNA (**Figure 2.3.**) [17]. Reverse transcription is accomplished as a result of the in-built polymerase-dependent activity in RT, such as the RNA-dependent DNA polymerase, RNase-H and DNA-dependent DNA polymerase activities [17]. RT was the very first HIV-1 protein to be used as a target in anti-HIV drug development [17].

As illustrated in **Figure 2.3.**, RT is an asymmetrical heterodimer comprised of a 66-kDa subunit and a 51-kDa subunit with 560 and 460 amino acid residues, respectively and is connected to each other by a common amino terminus [39, 40]. The p66 subunit (**Figure 2.3.**) is further divided

into the polymerase and RNase H domains, which are spatially distinct in nature [40]. The structure of the polymerase domain (**Figure 2.3.**) resembles that of a right hand, however, is comprised of three highly conserved regions, and i.e. fingers domain, palm domain and thumb domain, respectively [41, 42]. Furthermore, the highly active RNase H domain and connection domain are located in this p66 subunit (**Figure 2.3.**). Although, the p51 subunit consists of the identical domains situated in the p66 subunit, however, these are orientated differently [40]. These different positions are the main reason that the p51 subunit folds differently from the p66 subunit. Even though the p51 subunit lacks enzymatic activity, it stills provides a supportive framework for p66 catalytic activities [43-45].
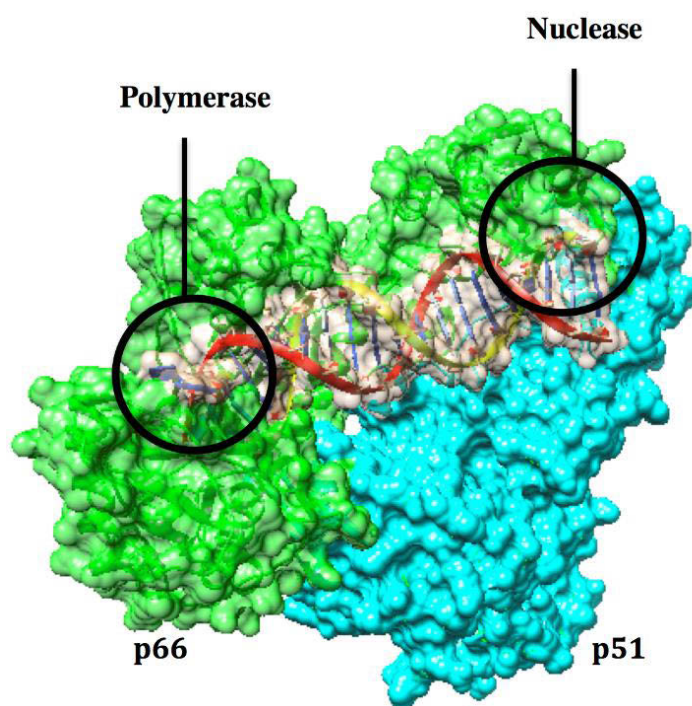
Figure 2.3. The X-ray crystal structure of reverse transcriptase (RT) (PDB code: 3KLF) with the P51 subunit (cyan) and the P66 subunit (green) respectively [40].

Reverse transcriptase is a key role player in several distinct biological functions all of which are essential in the retrotranscription procedure. These include: RNase H catalysis, RNA-and-DNA-dependent DNA synthesis, strand transfer and RNA cleavage [17]. The highly dynamic nature of RT is primarily responsible for the execution of its array of functions. It is this very high dynamic nature that permits the unrestrained gliding of RT over long distances of RNA/DNA and DNA/DNA complexes, to effortlessly target the primer terminus for DNA polymerization,

and to gain quick access to numerous sites. It was recently reported that RT has an interchangeable dual binding mode of action, enabling RT to bind to different nucleic acid substrates [39]. These two different binding positions are termed RNase H cleavage component orientation" and "polymerase component orientation" and co-exist in dynamic equilibrium.

Reverse transcriptase is the target for two distinct classes of antiviral drugs, nucleoside RT inhibitors (NRIs) and non-nucleoside RT inhibitors (NNRTIs) [46], which will be discussed in detail in the following sections.

2.4.2. Integrase (IN)

HIV-1 integrase (IN) **(Figure 2.4.)** was one of the most recent HIV-1 enzymes to be featured in antiretroviral drug discovery [47, 48]. The highly replicative and infectiveness nature of HIV-1 is predominantly dependent upon the insertion of a single DNA copy enclosed in the viral genome into the host cell chromosome in a process called integration [49]. The HIV-1 IN catalyzes the various DNA cleaving and ligating steps that constitute the integration of HIV-1 into the host cell genome [50-54]. More specifically, IN catalyzes endonucleolytic processing of the 3' end of dsDNA and DNA strand transfer [53, 54].

The HIV-1 IN enzyme is 31-kDa in size and composed of 288 amino acid residues and three distinct domains: an N-terminal domain, the catalytic core and an C-terminal domain, respectively **(Figure 2.4.)** [37, 49, 55]. It has been reported that all three domains are indispensable to integration, however, it is postulated that the catalytic domain contains principle active site amino acids, which are essential for catalyzing both the chemical reactions of integration and disintegration processes [56-59]. The N-terminal encloses a His_2Cyst_2 zinc-binding motif **(Figure 2.4.)**, which indicates that this domain interacts with nucleic acids; whereas, the C-terminal domain houses a DNA binding domain **(Figure 2.4.)** indicating that this domain can facilitate binding to both viral and host genome [53, 54].
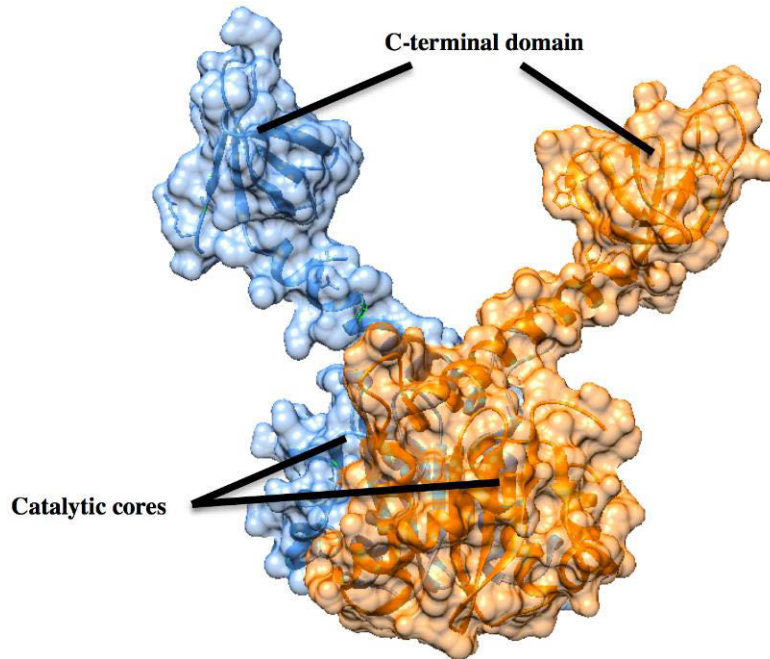
Figure 2.4. Structure of HIV-1 integrase (IN) enzyme (PBD code: 1EX4) with the N-terminal (HH-CC zinc-binding catalytic core) domain and C-terminal (DNA binding domain) sections.

HIV-1 IN catalyzes 3' end processing and DNA strand transfer [47, 48, 60]. Since integration positions in the host genome are non-specific, assimilation can occur at any possible site. Therefore, HIV-1 IN has become a promising target in anti-HIV drug discovery. Thus, IN inhibitors (INIs/InSTIs) to be discussed in the following section are highly specific, as these drugs target and inhibit strand transfer and hence block the insertion of viral DNA into cellular DNA [61-64]. As the integration of HIV-1 genetic material (i.e. RNA) into the host DNA is crucial for preserving the dsDNA found in an infected cell as well as is important for viral protein and RNA expression. The mechanism mentioned previously is unique to HIV-1 and is therefore regarded as a favorable therapeutic target.

2.4.3. Protease (PR)

The HIV-1 protease (PR) enzyme has been investigated as a strategic target in anti-HIV drug discovery (**Figure 2.5.**) [65]. As stated earlier, HIV-1 PR is responsible for the catalytic cleavage of newly synthesized polypeptide precursor proteins, *gag* and *gag-pol,* which are essential elements for virion maturation [66-69].

Investigation of the X-ray crystal structure of HIV-1 PR revealed that PR is a dimeric protein, with each component containing 99 amino acid residues and is 11-kDa in size [70, 71]. As shown in **Figure 2.5**, HIV-1 PR is composed of three canonical domains: N-terminal domain (active site cavity), middle domain and C-terminal domain (dimerization cavity). Also, a flexible flap region domain encloses the active site and does so more during substrate binding (**Figure 2.5.**) [34, 72]. It was reported that the active site cavity located in the N-terminal domain is fundamental in stabilization of the dimer and catalytic sites [1, 34, 70].
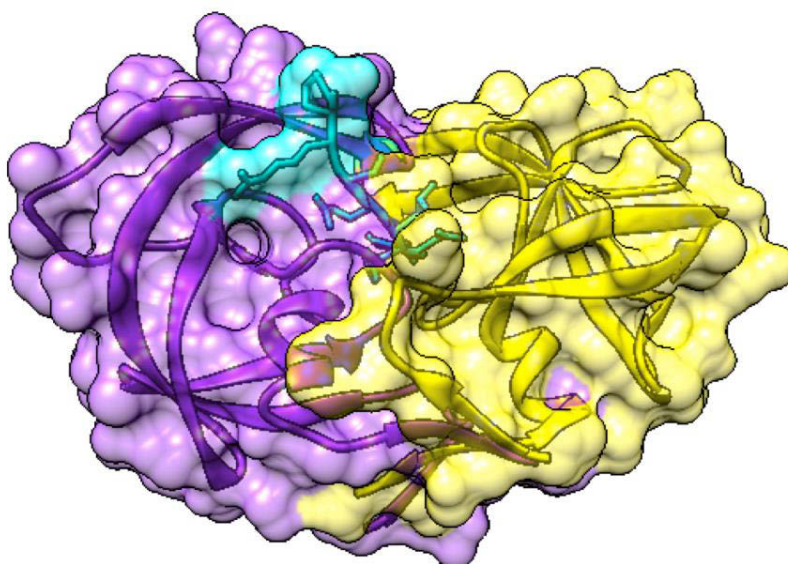
Figure 2.5. Structure of the dimer of HIV-1 protease (PR) (PDB code: 3QOZ) with the two-dimerization domains (purple and yellow), flaps, a polypeptide substrate (cyan) and active site domain (Asp 25 amino acids –cyan)

As mentioned above, HIV-1 PR is an integral enzyme in catalytic splicing of viral polypeptide precursor proteins that are required for virion maturation. Therefore, HIV-1 PR presents itself as an attractive target in antiretroviral drug development. Protease inhibitors (PIs) bind and inhibit PR enzyme, and thus the HIV-1 life cycle is interrupted as the virus is prohibited from maturing and therefore cannot transform into its infectious stage. Protease inhibitors will further be discussed in the following sections.

To date, there are currently several FDA-approved drugs in clinical use and commercially available for HIV/AIDS therapeutic regime.

2.4.4. New avenues for the discovery of potential anti-HIV targets

Numerous concerns regarding the long-term side effects of antiretroviral drugs and the increasing transmission of resistant variants accentuates the requirement to identify new classes of drugs, which are able to suppress HIV-1 replication efficiently [9]. Thus, there is an on-going need for novel therapeutics, which can prevent the entry of HIV-1 into its target cells [73, 74]. Therefore, the next section provides insight into the structural and functional properties of two such new avenue structures, co-receptor chemokine receptor 5 (CCR5) and negative factor (*Nef*) as potential anti-HIV targets.

2.4.4.1. Chemokine Co-receptor receptor 5 (CCR5)

The entry of the HIV virus into its target cell is mediated by the specific interactions of the target cell itself such as the interaction between gp120 viral envelope glycoprotein and the plasmatic membrane receptors [75].

Besides CD4, certain chemokine receptors (CCRs) belonging to the G-protein coupled receptor superfamily (GPCRs) have recently been identified as co-targets essential for HIV viral entry into target cells [73, 76]. The chemokine co-receptor 5 (CCR5) (**Figure 2.6.**) has been reported to be a crucial pharmaceutical target in HIV-1 infection and other inflammatory diseases. CCR5 acts as an integral co-receptor to gp120 glycoprotein during HIV-1 cell membrane attachment.
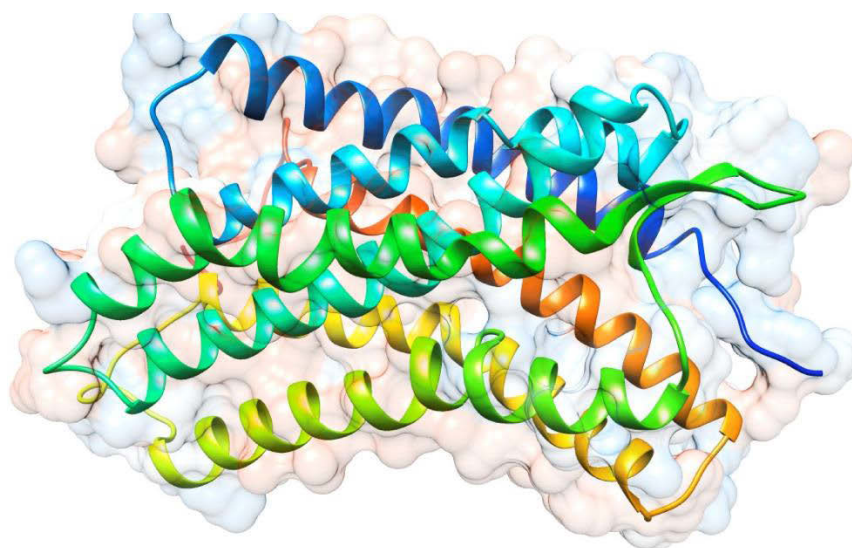
Figure 2.6. Structure of CCR5 co-receptor protein (PDB code: 3ODU).

To date, limited research that has been conducted on this CCR5 co-receptor protein (**Figure 2.6.**) [77-80]. Despite the fact that no structural information exists about CCR5 precise binding site and lack of an X-ray crystal structure, few studies have investigated and reported findings on creating homology models and designing small molecule CCR5 antagonists [77-80]. Homology modeling of CCR5 (**Figure 2.6.**) has been reported before, where CXCR4 was used as a structural template [73].

2.4.4.2. Negative factor (*Nef*)

In addition to the reported structural proteins, HIV-1 is responsible for encoding of four different accessory proteins, which are required for immune evasion and optimization for viral replication. [81, 82]. Among these, is HIV-negative factor (*Nef*) (**Figure 2.7.**), a small 27-35 kDa myristoylated protein, which has an imperative role in HIV-1 pathogenicity and more specifically expressed as a virulence factor [81, 82]. Nef is produced in large quantities during the early stages of viral gene expression [7].

Nef is a myriostoylated protein that is composed of 206 amino acid residues and is located in the cytoplasm as well as has association with cellular membranes (**Figure 2.7.**) [83]. The structure of HIV-1 Nef (**Figure 2.7.**) is characterized by its flexible loop regions that contain several sequence motifs as an N-terminal myristoylation site, a central poly-proline PxxP motif for SH3 domain binding and C-terminal motifs for interaction with clathrin- associated endosomal adaptor protein complexes [83]. As illustrated in **Figure 2.7**, Nef is composed of two dimers, and with each dimer been composed of a large sub-domain and small sub-domain and these two dimers subunits undergo dimerization in order for Nef protein to become activated, manipulate host's cellular machinery and allow for pathogenic activity in HIV replication [81, 82].
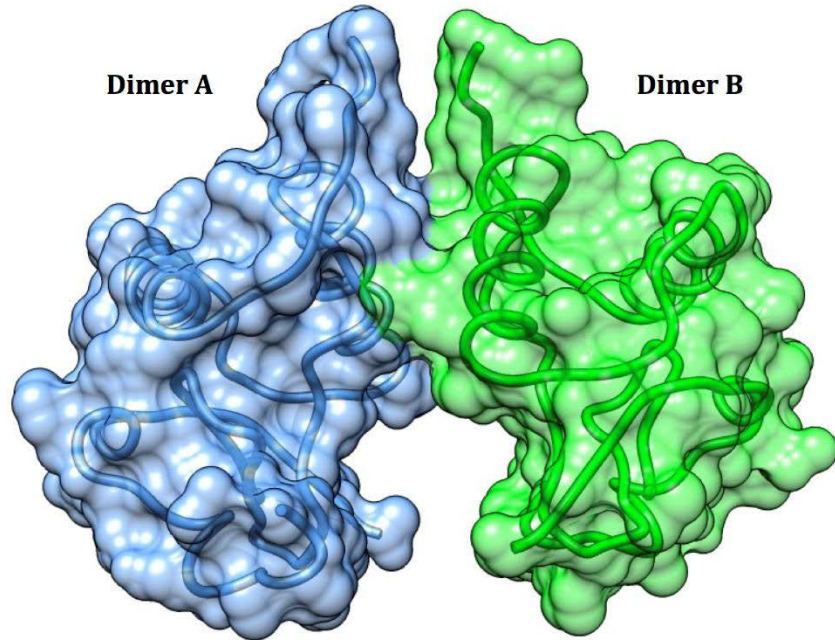
Figure 2.7. Illustration of the HIV-1 Nef dimer protein (PDB code: 1EFN) with two flexible helical loops.

Because of its major role in HIV-1 pathogenicity, Nef protein has been proved to be a very important target in anti-HIV drug design and discovery process

2.5. Antiretroviral therapy (HAART)

In AIDS therapy, the fundamental strategy is to inhibit viral replication and hence, to slow down the destruction of the immune system and prolong the lives of infected individuals. Despite the ongoing HIV/AIDS research, to date, no cure has yet been found for this disease [84]. Currently, the most effective treatment comprises a complex "cocktail" FDA-approved and clinical trial drugs that involve multiple drug targets [85-88]. These drugs target and attack the virus at different stages of its lifecycle, thereby halting virus replication and reducing destruction of the immune system. These include: protease inhibitors (PIs), reverse transcriptase (RT) inhibitors, integrase (IN) inhibitors and entry inhibitors [89]. The first antiretroviral drug to be approved by the FDA in 1987 was a nucleoside reverse transcriptase inhibitor (NRTI) called Zidovudine (AZT) [90]. Similarly, in 1991, another NRTI, Zalcitibine, was approved for clinical use [91]. Despite the fact with the first FDA-approved protease inhibitor, Saquinavir identified in 1995, to date, only nine FDA-approved PIs exists [92]. Likewise, in 2007, Raltegravir [93], the first integrase inhibitor was

approved by the FDA followed by the second INI, Elvitegravir [94] been approved in 2012. The latest addition to the class of INIs was Doultegravir, which was approved in 2013 [95].

Numerous concerns regarding the long-term side effects of antiretroviral drugs and the increasing transmission of resistant variants accentuates the requirement to identify new classes of drugs, which are able to suppress HIV-1 replication efficiently [9]. Therefore, there is an on-going need for novel therapeutics, which can target and inhibit the virus at various stages of its life cycle [73, 74].

2.5.1. Reverse transcriptase inhibitors

Since HIV-1 RT is a key role player in viral replication, thus RT has become a significant target in drug therapy [96]. These inhibitors mainly interfere with the reverse transcription of viral RNA into DNA. In this section, the two distinct classes of approved RT inhibitors (**Figure 2.8.**): nucleoside analog RT inhibitors (NRTIs) [97, 98] and non-nucleoside RT inhibitors (NNRTIs) [97, 99] are explained in terms of their structural and mechanistic action features. NRTIs are analogs of native nucleoside substrates, whereas NNRTIs bind to a non-catalytic allosteric pocket on RT. The 12 FDA-approved agents as shown in **Figure 2.8** combined from both these classes' accounts for almost half of all approved antiretroviral drugs [92].
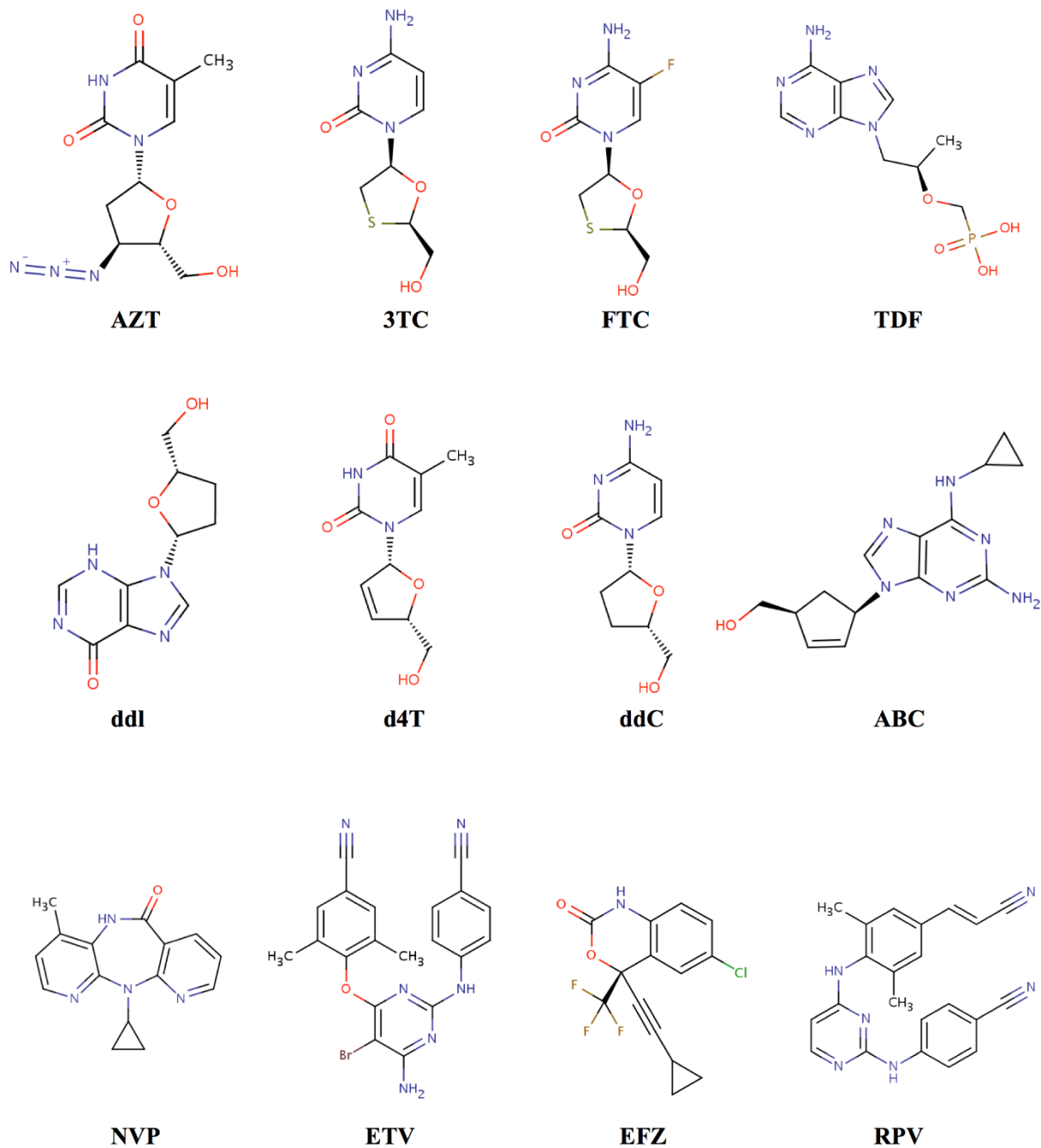
Figure 2.8. 2D representation of the 12 FDA-approved RT inhibitors: nucleoside analog RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs).

2.5.1.1. Nucleoside reverse transcriptase inhibitors (NRTIs)

As mentioned previously, AZT, an NRTI (**Figure 2.8.**), was the first reported antiretroviral drug to be approved in 1987 for HIV treatment [90]. NRTIs are different from other natural

deoxynucleotides as these RT inhibitors lack the 3'-hydroxyl group found on the deoxyribose moiety [88.] NRTIs compete with natural nucleoside substrates for binding to the RT polymerase active site, become incorporated into the primer site, covalently bind to DNA and then terminate proviral DNA synthesis [97, 99]. When NRTIs become incorporated during DNA transcription, nucleosides cannot be covalently added and this prevents 5'-3' phosphordiester bond formation, and thus transcription is terminated [100]. There are currently eight FDA-approved NRTIs as depicted in **Figure 2.8.**, and these include: zidovudine (AZT), lamivudine (3TC), emtricitabine (FTC), tenofovir (TDF), didanosine (ddl), stavudine (d4T), zalcitabine (ddC) and abacavir (ABC) [92 100].

The next section will discuss the other second type of RT inhibitors, non-nucleoside reverse transcriptase inhibitors (NNRTIs), which exhibits less toxicity compared to NRTIs.

2.5.1.2. Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) (**Figure 2.8.**) were first reported in 1990. This class of RT inhibitors have more site-specific interactions with the allosteric (non-substrate binding) site on HIV-RT. When RT active site undergoes conformational changes, it prevents viral replicative process [101]. Currently, there are five approved NNRTIs (**Figure 2.8.**): nevirapine (NVP), etravirine (ETV), efavirenz (EFV) and rilpivirine (RPV) reported during 1996-2011 [101 97, 99].

2.5.2. Integrase inhibitors

Integrase was one of the most recent HIV-1 enzyme to have been successfully targeted for anti-HIV therapy [47, 48]. During the past two decades, a variety of small candidate molecule inhibitors have been identified from high-throughput screening and other diverse chemical classes as potential integrase inhibitors (INIs) [61-64]. Diketo aryl (DKA) and DKA-like inhibitors, or which are more commonly known as strand-transfer inhibitors have been identified to date as the most promising INIs or INSTIs. These inhibitors either halts HIV transcription by blocking the DNA strand transfer reaction or by inhibiting the Tat-TAR interaction [61-64]. Raltegravir, was the first integrase inhibitor to be approved by the FDA in 2007 [93], followed by elvitegravir approved in 2012 [94] and dolutegravir approved in 2013 [95].

Raltegravir   **Dolutegravir**
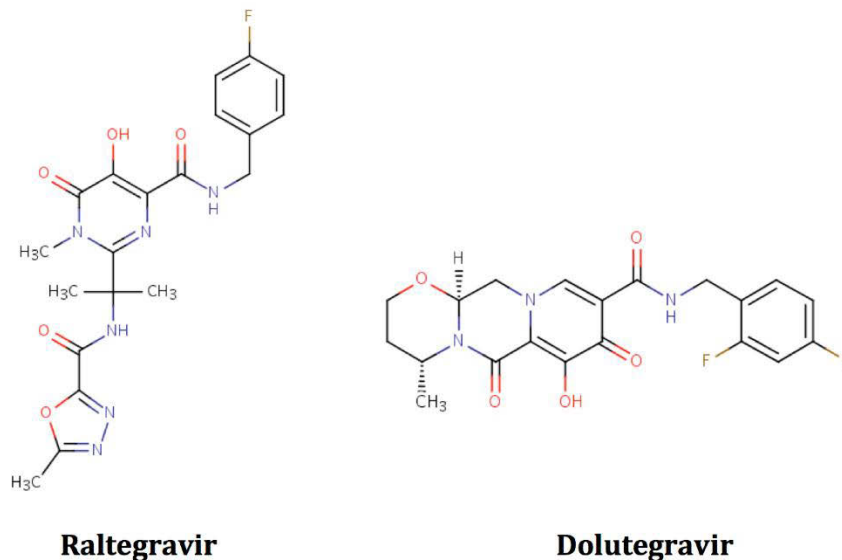
Figure 2.9. 2D representation of two FDA-approved Integrase inhibitors (INIs): Raltegravir and Dolutegravir, respectively.

2.5.3. Protease Inhibitors

As stated earlier, HIV-1 protease is responsible for the catalytic cleavage of newly formed polypeptide protein precursors into smaller particles required for virion maturation [34]. Therefore, the timing step of this stage is very critical and thus presents HIV-1 PR as a highly favorable target for drug therapy. Protease inhibitors (**Figure 2.10.**) are the class of antiretroviral drugs that target and bind to HIV PR and therefore prevent the virus from maturing and proceeding into its infectious stage [1, 71, 103, 104]. Protease inhibitors (**Figure 2.10.**) are considered amongst some of the most remarkable achievements in modern medicine. The mechanism of action of PR inhibitors is to inhibit the catalytic activity of PR enzyme, which disrupts new virion maturation, and thus renders virions into a non-infectious state [1, 71, 103, 104].
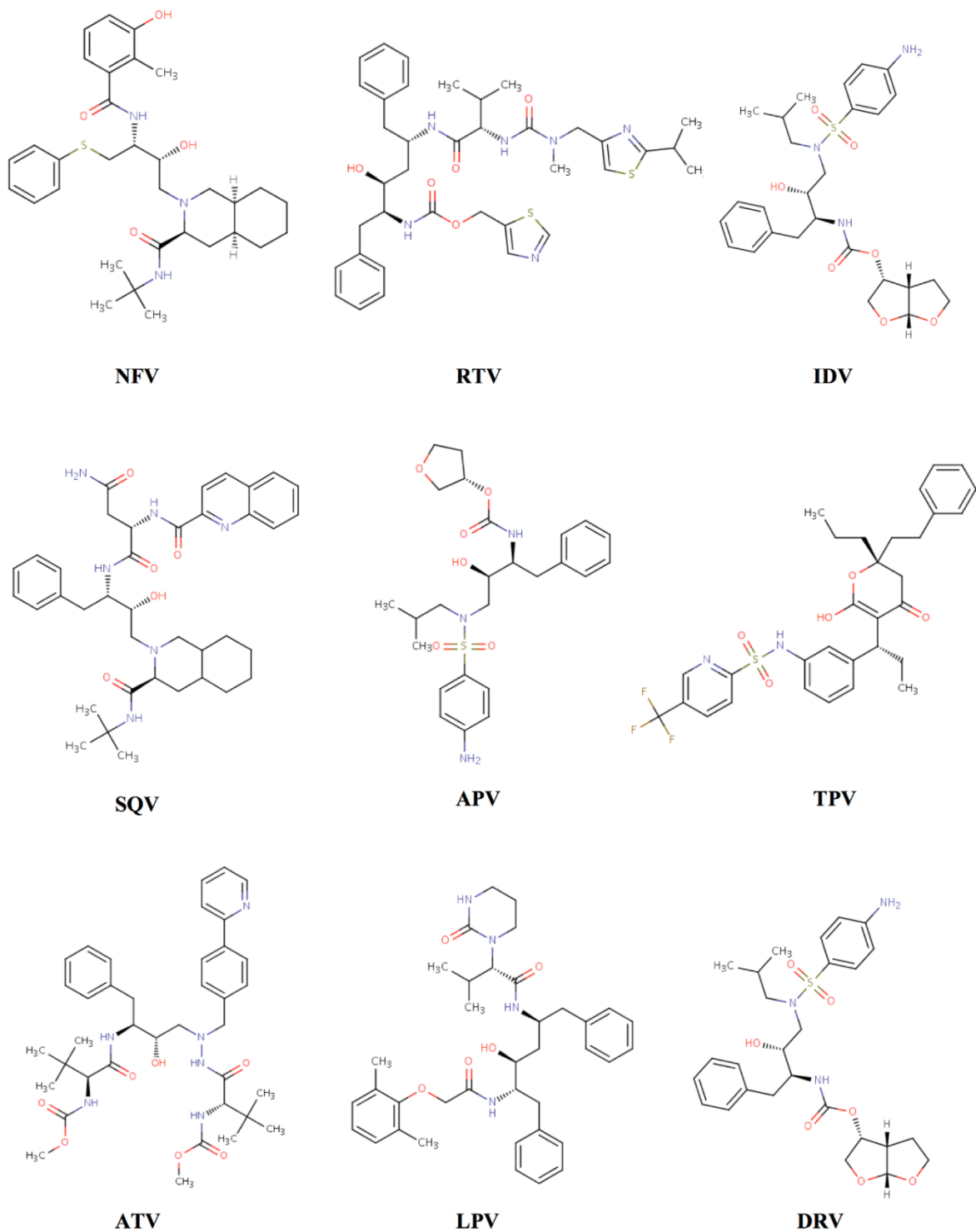
Figure 2.10. 2D representation of FDA-approved anti-HIV PR drugs.

Since the introduction of HIV PR inhibitors (**Figure 2.10.**) in 1995 [92], these anti-HIV drugs have been widely used, more often as the basis in HAART, i.e. the combination of PIs with other anti-

HIV drugs [103, 104]. It has been reported that the introduction of HAART has resulted in decrease mortality and prolonged life expectancy of numerous HIV infected individuals. As shown in **Figure 2.10.**, there are currently nine PIs approved by the FDA: nelfinavir (NFV), ritonavir (RTV), indinavir (IDV), saquinavir (SQV), amprenavir (APV), tipranavir (TPV), atazanavir (ATV), lopinavir (LPV) and darunavir (DRV) [105].

2.5.4. Entry inhibitors

These are competitive type of inhibitors that are specific, because these inhibitors have identical chemical and molecular geometrical resemblance to the substrate and competes with the substrate for same active site. There are three types of entry inhibitors: fusion inhibitors, attachment and co-receptor antagonist inhibitors. However, only the latter type of entry inhibitors, co-receptor antagonists' inhibitors will be described in terms of an example and mechanism.

2.5.4.1. Co-receptor antagonists' inhibitors

As mentioned previously in **section 2.4.4.1,** there are two types of co-receptors namely, CCR5 and CXCR4 that are expressed on T-cell surface. Co-receptor antagonists were the first class of anti-HIV drugs, which did not target the HIV virus directly, however, these drugs bind, fuse and block viral attachment to the host cell when they prevent virus/co-receptor interaction. Maraviroc, as depicted in **Figure 2.11**, was the first entry inhibitor approved as a CCR5 co-receptor antagonist in 2007 [106-108].
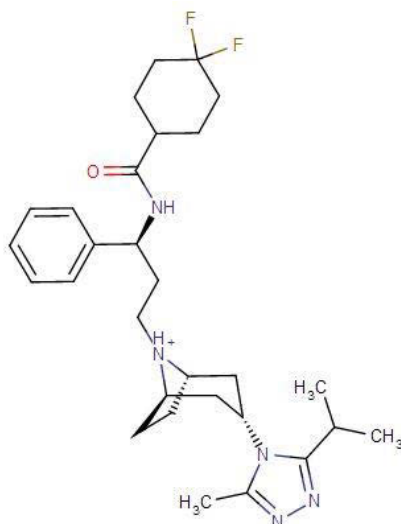
Figure 2.11. Maraviroc, the first FDA-approved CCR5 co-receptor antagonist inhibitor.

2.5.5. HIV accessory negative factor (*Nef*) inhibitors

Considerate insight into the dimerization and inhibition mechanism of HIV-1 accessory negative factor, *Nef*, is critical in the drug design process of potential *Nef* antagonists. Thus far, only a handful of small molecule inhibitors identified *via* high throughput screening (HTS) blocks *Nef*-dependent viral replicative process. One such molecule identified is a diphenylpyrazolodiazene containing small molecule inhibitor, B9 (**Figure 2.12.**), binds at the dimeric interface of HIV-*Nef* and prevents dimerization of the two *Nef* subunits [109, 110].
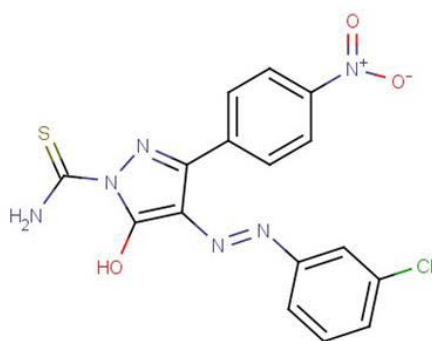
Figure 2.12. The 2D chemical structure of the only known Nef inhibitor, B9.

2.5.6. References

1. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.

2. Merson, M. H., O'Malley J Fau - Serwadda, D., Serwadda D Fau - Apisuk, C., and Apisuk, C. (2008) The history and challenge of HIV prevention, *Lancet Infectious Diseases 372*, 497-500.

3. UNAIDS. (2011) World AIDS day report 2011, *http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/jc2216_worldaidsday_report_2011_en.pdf*, 11 Nov. 2013.

4. UNAIDS. (2012) UNAIDS report on the global AIDS epidemic 2012, *http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf*, 20 Jan. 2014.

5. UNAIDS. (2012) Global summary of the AIDS epidemic, *http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/201309_epi_core_en.pdf*, 20 Jan 2014.

6. Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Medicinal Chemistry Research*.

7. Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. . (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants, *Retrovirology* 1-23.

8. 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 & Drug Design 81*, 208-218.

9. Xu, Y., Liu, H., Niu, C. Y., Luo, C., Luo, X. M., Shen, J. H., Chen, K. X., and Jiang, H. L. (2004) Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-

yl)-butanes based on the structural modeling of human CCR5 receptor, *Bioorganic & Medicinal Chemistry 12*, 6193-6208.

10. Taylor, J. E., and Korber, B. T. (2005) HIV-1 intra-subtype superinfection rates: estimates using a structured coalescent with recombination, *Infection, Genetics and Evolution 5*, 85-95.

11. Clavel, F., Guetard, D., Brunvezinet, F., Chamaret, S., Rey, M. A., Santosferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L., and Montagnier, L. (1986) Isolation of a new human retrovirus from west-african patients with AIDS, *Science 233*, 343-346.

12. Gallo, S. A., Reeves, J. D., Garg, H., Foley, B., Doms, R. W., and Blumenthal, R. (2006) Kinetic studies of HIV-1 and HIV-2 envelope glycoprotein-mediated fusion, *Retrovirology 3*.

13. McCutchan, F. E. (2006) Global epidemiology of HIV, *Journal of Medical Virolgy 78*, S7-S12.

14. Sharp, P. M., Bailes, E., Robertson, D. L., Gao, F., and Hahn, B. H. (1999) Origins and evolution of AIDS viruses, *The Biological Bulletin 196*, 338-342.

15. de Silva, T. I., Cotten, M., and Rowland-Jones, S. L. (2008) HIV-2: the forgotten AIDS virus, *Trends Microbiol. 16*, 588-595.

16. Reeves, J. D., and Doms, R. W. (2002) Human immunodeficiency virus type 2, *Journal of General Virology 83*, 1253-1265.

17. Hu, W.-S., and Hughes, S. H. (2012) HIV-1 reverse transcription, *Cold Spring Harbor perspectives in medicine 2*.

18. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV, *Journal of Molecular Biology 285*, 1-32.

19. Levy, J. A. (2007) *HIV and the Pathogenesis of AIDS, 3rd Edition, American Society of Microbiology*, 1752 N Street Nw, Washington, Dc 20036-2904 USA.

20. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein, *Cell 89*, 263-273.

21. McKeage, K., and Perry, C. M. (2002) Trastuzumab - A review of its use in the treatment of metastatic breast cancer overexpressing HER2, *Drugs 62*, 209-243.

22. Reinhard Renneberg, A. L. D. (2008) Biotechnology for beginners, Springer, Germany.

23. Cunningham, A. L., Donaghy, H., Harman, A. N., Kim, M., and Turville, S. G. (2010) Manipulation of dendritic cell function by viruses, *Current Opinion in Microbiology 13*, 524-529.

24. Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) The CD4(T4) antigen is an essential component of the receptor for the AIDS retrovirus, *Nature 312*, 763-767.

25. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., and Montagnier, L. (1984) Lymphocyte-T T4 molecule behaves as the receptor for human retrovirus LAV, *Nature 312*, 767-768.

26. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease, *Annual Review of Immunology 17*, 657-700.

27. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature 393*, 648-659.

28. Levy, J. A. (2009) HIV pathogenesis: 25 years of progress and persistent challenges, *Aids 23*, 147-160.

29. Alkhatib, G., Combadiere, C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M., and Berger, E. A. (1996) CC CKRS: A RANTES, MIP-1 alpha, MIP-1 beta receptor as a fusion cofactor for macrophage-tropic HIV-1, *Science 272*, 1955-1958.

30. Craigie, R., and Bushman, F. D. (2012) HIV DNA Integration, *Cold Spring Harbor perspectives in medicine 2*, a006890.

31. Zheng, Y. H., Lovsin, N., and Peterlin, B. M. A. (2005) Newly identified host factors modulate HIV replication, *Immunology Letters 97*, 225-234.

32. Karn, J., and Stoltzfus, C. M. (2012) Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression, *Cold Spring Harbor perspectives in medicine 2*, a006916.

33. Spearman, P. (2006) Cellular cofactors involved in HIV assembly and budding, *Current Opinion in HIV and AIDS 1*, 200-207.

34. Brik, A., and Wong, C. H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & Biomolecular Chemistry 1*, 5-14.

35. Van Lint, C., Bouchat, S., and Marcello, A. (2013) HIV-1 transcription and latency: an update, *Retrovirology 10*.

36. Esposito, F., Corona, A., and Tramontano, E. (2012) HIV-1 reverse transcriptase still remains a new drug target: structure, function, classical inhibitors, and new inhibitors with innovative mechanisms of actions, *Molecular Biology International 2012*, 1-23.

37. Vangent, D. C., Vink, C., Groeneger, A., and Plasterk, R. H. A. (1993) Complementation between HIV integrase proteins mutated in different domains, *EMBO Journal 12*, 3261-3267.

38. Tronchet, J. M. J., and Seman, M. (2003) Nonnucleoside inhibitors of HIV-1 reverse transcriptase: From the biology of reverse transcription to molecular design, *Current Topics in Medicinal Chemistry 3*, 1496-1511.

39. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor, *Science 256*, 1783- 1790.

40. Rigby, M. (1986) Physical chemistry, 3rd edition - ATKINS,PW, *Nature 319*, 820-820.

41. Ilina, T., LaBarge, K., Sarafianos, S.G., Ishima, R., and Parniak, M.A. (2012) Inhibitors of HIV-1 Reverse TranscriptaseAssociated Ribonuclease H Activity., *Biology 1*, 521-541.

42. Harpstead, B. (2005) HIV-1 reverse transcriptase structure, activities, and inhibitors., *Basic Biotechnology MMG 445*, 1-7.

43. Bruccoleri, A. (2013) Positional Adaptability in the Design of Mutation-Resistant Nonnucleoside HIV-1 Reverse Transcriptase Inhibitors: A Supramolecular Perspective., *AIDS Research and Human Retroviruses 29*, 1-9.

44. Esposito, F., Corona, A., and Tramontano, E. (2012) HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with InnovativeMechanisms of Actions., *Molecular Biology International 2012*, 1-23.

45. Seckler, J. M., Howard, K.J., Barkley, M.D., and Wintrode, P.L. (2009) Solution Structural Dynamics of HIV-1 Reverse Transcriptase Heterodimer., *Biochemistry 48*, 7646–7655.

46.   Young, F. E. (1988) The role of the FDA in the effort against AIDS, *Public Health Reports 103*, 242-245.

47.   Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J. Y., Young, S., Hamill, T., Cole, J. L., and Hazuda, D. J. (2000) HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase, *Proceedings of the National Academy of Sciences U. S. A. 97*, 11244-11249.

48.   Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L. H., Fisher, T. E., Embrey, M., Guare, J. P., Egbertson, M. S., Vacca, J. P., Huff, J. R., Felock, P. J., Witmer, M. V., Stillmock, K. A., Danovich, R., Grobler, J., Miller, M. D., Espeseth, A. S., Jin, L. X., Chen, I. W., Lin, J. H., Kassahun, K., Ellis, J. D., Wong, B. K., Xu, W., Pearson, P. G., Schleif, W. A., Cortese, R., Emini, E., Summa, V., Holloway, M. K., and Young, S. D. (2004) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase, *Proceedings of the National Academy of Sciences U. S. A. 101*, 11233-11238.

49.   Engelman, A., Bushman, F. D., and Craigie, R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex, *EMBO Journal 12*, 3269-3275.

50.   Chiu, T. K., and Davies, D. R. (2004) Structure and function of HIV-1 integrase, *Current Topics in Medicinal Chemistry 4*, 965-977.

51.   Craigie, R. (2001) HIV integrase, a brief overview from chemistry to therapeutics, *Journal of Biological Chemistry 276*, 23213-23216.

52.   Di Santo, R., Costi, R., Artico, M., Tramontano, E., La Colla, P., and Pani, A. (2003) HIV-1 integrase inhibitors that block HIV-1 replication in infected cells. Planning synthetic derivatives from natural products, *Pure and Applied Chemistry 75*, 195-206.

53.   Brin, E., Yi, J., Skalka, A. M., and Leis, J. (2000) Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration, *Journal of Biological Chemistry 275*, 39287-39295.

54.   Yoder, K. E., and Bushman, F. D. (2000) Repair of gaps in retroviral DNA integration intermediates, *Journal of Virology 74*, 11191-11200.

55. Engelman, A., and Craigie, R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro, *Journal of Virology 66*, 6361-6369.

56. Barreca, M. L., Ferro, S., Rao, A., De Luca, L., Zappala, M., Monforte, A. M., Debyser, Z., Witvrouw, M., and Chimirri, A. (2005) Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors, *Journal of Medicinal Chemistry 48*, 7084-7088.

57. d'Angelo, J., Mouscadet, J. F., Desmaële, D., Zouhiri, F., and Leh, H. (2001) HIV-1 integrase: the next target for AIDS therapy?, *Pathologie Biologie 49*, 237-246.

58. Grobler, J. A., Stillmock, K., Hu, B. H., Witmer, M., Felock, P., Espeseth, A. S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J. S., Young, S., Vacca, J., and Hazuda, D. J. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes, *Proceedings of the National Academy of Sciences of the United States of America 99*, 6661-6666.

59. Zhuang, L. C., Wai, J. S., Embrey, M. W., Fisher, T. E., Egbertson, M. S., Payne, L. S., Guare, J. P., Vacca, J. P., Hazuda, D. J., Felock, P. J., Wolfe, A. L., Stillmock, K. A., Witmer, M. V., Moyer, G., Schleif, W. A., Gabryelski, L. J., Leonard, Y. M., Lynch, J. J., Michelson, S. R., and Young, S. D. (2003) Design and synthesis of 8-hydroxy- 1,6 naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells, *Journal of Medicinal Chemistry 46*, 453-456.

60. Carson, M. (1987) Ribbon models of macromolecules, *Journal of Molecular Graphics 5*, 103-106.

61. Hsu, M. C., Schutt, A. D., Holly, M., Slice, L. W., Sherman, M. I., Richman, D. D., Potash, M. J., and Volsky, D. J. (1991) Inhibition of HIV replication in acute and chronic infections invitro by a tat antagonist, *Science 254*, 1799-1802.

62. Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., AboulEla, F., Varani, G., Karn, J., and Klimkait, T. (1997) An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication, *Proceedings of the National Academy of Sciences U. S. A. 94*, 3548-3553.

63. Cupelli, L. A., and Hsu, M. C. (1995) The human-immunodeficiency-virus type-1 tat antagonist, RO-5-3335, predominantly inhibits transcription initiation from the viral promoter, *Journal of Virology 69*, 2640-2643.

64. Hwang, S., Tamilarasu, N., Kibler, K., Cao, H., Ali, A., Ping, Y. H., Jeang, K. T., and Rana, T. M. (2003) Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication, *Journal of Biological Chemistry 278*, 39092-39103.

65. Wensing, A. M. J., van Maarseveen, N. M., and Nijhuis, M. (2010) Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance, *Antiviral Research 85*, 59-74.

66. Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C. T., Lumma, P. K., Freidinger, R. M., Veber, D. F., and Sigal, I. S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins, *Biochemical and Biophysical Research Commununity 156*, 297-303.

67. Katz, R. A., and Skalka, A. M. (1994) The retroviral enzymes, *Annual Review of Biochemistry 63*, 133-173.

68. Mager, P. P. (2001) The active site of HIV-1 protease, *Medical Research Reviews 21*, 348-353.

69. Sule, P., and Nagy, A. (1996) Density functional study of strong hydrogen-bonded systems: The hydrogen diformiate complex, *The Journal of Chemical Physics 104*, 8524-8534.

70. Castro, H. C., Abreu, P. A., Geraldo, R. B., Martins, R. C. A., dos Santos, R., Loureiro, N. I. V., Cabral, L. M., and Rodrigues, C. R. (2011) Looking at the proteases from a simple perspective, *Journal of Molecular Recognition 24*, 165-181.

71. Martinez-Cajas, J. L., and Wainberg, M. A. (2007) Protease inhibitor resistance in HIV-infected patients: Molecular and clinical perspectives, *Antiviral Research 76*, 203-221.

72. Weber, I. T., and Agniswamy, J. (2009) HIV-1 protease: Structural perspectives on drug resistance, *Viruses-Basel 1*, 1110-1136.

73. Fano, A., Ritchie, D. W., and Carrieri, A. (2006) Modeling the structural basis of human CCR5 chemokine receptor function: From homology model building and molecular dynamics validation to agonist and antagonist docking, *Journal of Chemical Information and Modeling 46*, 1223-1235.

74. Manikandan, S., and Malik, B. K. (2008) Modeling of human CCR5 as target for HIV-I and virtual screening with marine therapeutic compounds, *Bioinformation 3*, 89-94.

75.   Cormier, E. G., Persuh, M., Thompson, D. A. D., Lin, S. W., Sakmar, T. P., Olson, W. C., and Dragic, T. (2000) Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120, *Proceedings of the National Academy of Sciences 97*, 5762-5767.

76.   F, C., AL, D., A, G.-D., SK, A., RC, G., and P, L. - Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major, *D - 0404511*, T - ppublish.

77.   Perez-Nueno, V. I., Ritchie, D. W., Rabal, O., Pascual, R., Borrell, J. I., and Teixido, J. (2008) Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking, *Journal of Chemical Information and Modeling 48*, 509-533.

78.   Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., and Igglessi-Markopoulou, O. (2006) Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques, *Journal of Computer-Aided Molecular Design 20*, 83-95.

79.   Aher, Y. D., Agrawal, A., Bharatam, P. V., and Garg, P. (2007) 3D-QSAR studies of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas as CCR5 receptor antagonists, *Journal of Molecular Modeling 13*, 519-529.

80.   Kellenberger, E., Springael, J.-Y., Parmentier, M., Hachet-Haas, M., Galzi, J.-L., and Rognan, D. (2007) Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening, *Journal of Medicinal Chemistry 50*, 1294-1303.

81.   Sarafianos, S. G., Das, K., Hughes, S. H., and Arnold, E. (2004) Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases, *Current Opinion in Structural Biology 14*, 716-730.

82.   Das, S. R., and Jameel, S. (2005) Biology of the HIV Nef protein, *Indian Journal of Medical Research 121*, 315-332.

83.   Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996) Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain, *Cell 85*, 931-942.

84.   Cohen, M. S., Smith, M. K., Muessig, K. E., Hallett, T. B., Powers, K. A., and Kashuba, A. D. (2013) Antiretroviral treatment of HIV-1 prevents transmission of HIV-1: where do we go from here?, *The Lancet 382*, 1515-1524.

85.   Johnson, B. C., Metifiot, M., Ferris, A., Pommier, Y., and Hughes, S. H. (2013) A Homology Model of HIV-1 Integrase and Analysis of Mutations Designed to Test the Model, *Journal of molecular biology 425*, 2133-2146.

86.   Zhan, P., Chen, X., Li, D., Fang, Z., De Clercq, E., and Liu, X. (2013) HIV-1 NNRTIs: structural diversity, pharmacophore similarity, and impliations for drug design, *Medical Reserach Reviews 33*, E1-E72.

87.   Patel, J. R., and Prajapati, L. M. (2013) Predictive QSAR modeling on tetrahydropyrimidine-2-one derivatives as HIV-1 protease enzyme inhibitors, *Medicinal Chemistry Research 22*, 2795-2801.

88.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets: Infectious Disorders 2*, 17-32.

89.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets - Infectious Disorders 2*, 17-32.

90.   Marwick, C. (1987) AZT (zidovudine) just a step away from fda approval for aids therapy, *JAMA 257*, 1281-1282.

91.   Anastasi, J. K., and Rivera, J. L. (1991) AIDS drug update: DDI and DDC, *RN 54*, 41-43.

92.   James, J. S. (1995) Saquinavir (Invirase): first protease inhibitor approved-reimbursement, information hotline numbers, *AIDS treatment news*, 1-2.

93.   Hicks, C., and Gulick, R. M. (2009) Raltegravir: The first HIV type 1 integrase inhibitor, *Clinical Infectious Diseases 48*, 931-939.

94.   FDA.   (2012) http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/antiviraldrugsadvisorycommittee/ucm305852.pdf, 04 Mar. 2014.

95.   FDA. (2014) Antiretroviral drugs used in the treatment of HIV infection, *http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm* 02 Feb. 2014.

96.   Ilina, T., LaBarge, K., Sarafianos, S. G., Ishima, R., and Parniak, M. A. (2012) Inhibitors of HIV-1 reverse transcriptase-associated ribonuclease H activity, *Biology 1*, 521-541.

97.   Vivet-Boudou, V., Didierjean, J., Isel, C., and Marquet, R. (2006) Nucleoside and nucleotide inhibitors of HIV-1 replication, *Cellular and Molecular Life Sciences 63*, 163-186.

98.   Bauman, J. D., Das, K., Ho, W. C., Baweja, M., Himmel, D. M., Clark, A. D., Oren, D. A., Boyer, P. L., Hughes, S. H., Shatkin, A. J., and Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design, *Nucleic Acids Research 36*, 5083-5092.

99.   Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses *Antiviral Reserach 72*, 256-256.

100.   Williams, I. G. (2003) Enfuvirtide (Fuzeon (R)): The first fusion inhibitor, *International Journal of Clinical Practice 57*, 890-897.

101.   FDA. (2014) Antiretroviral drugs used in the treatment of HIV infection, *http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm*, 02 Feb. 2014.

102.   Pommier, Y., Johnson, A. A., and Marchand, C. (2005) Integrase inhibitors to treat HIV/AIDS, *Nature Reviews: Drug Discovery 4*, 236-248.

103.   Krohn, A., Redshaw, S., Ritchie, J. C., Graves, B. J., and Hatada, M. H. (1991) Novel binding of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere, *Journal of Medicinal Chemistry 34*, 3340-3342.

104.   Kim, E. E., Baker, C. T., Dwyer, M. D., Murcko, M. A., Rao, B. G., Tung, R. D., and Navia, M. A. (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme, *Journal of the American Chemical Society 117*, 1181-1182.

105.   Hughes, P. J., Cretton-Scott, E., Teague, A., and Wensel, T. M. (2011) Protease Inhibitors for Patients With HIV-1 Infection: A Comparative Overview, *P & T : a peer-reviewed journal for formulary management 36*, 332-345.

106.   Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank

3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Research 39*, D1035-D1041.

107.   Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Research 36*, D901-D906.

108.   Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Research 34*, D668-D672.

109.   Emert-Sedlak, L. A., Narute, P., Shu, S. T., Poe, J. A., Shi, H., Yanamala, N., Alvarado, J. J., Lazo, J. S., Yeh, J. I., Johnston, P. A., and Smithgall, T. E. (2013) Effector Kinase Coupling Enables High-Throughput Screens for Direct HIV-1 Nef Antagonists with Antiretroviral Activity, *Chemistry & Biology 20*, 82-91.

110.   Breuer, S., Schievink, S. I., Schulte, A., Blankenfeldt, W., Fackler, O. T., and Geyer, M. (2011) Molecular Design, Functional Characterization and Structural Basis of a Protein Inhibitor Against the HIV-1 Pathogenicity Factor Nef, *Plos One 6*, e20033.

CHAPTER 3

3. Introduction to computational chemistry

3.1. Introduction

Computational chemistry, also known as molecular modeling, may be defined as modeling of chemistry at atomic or molecular level using a set of techniques on a computer to investigate chemical problems. The term encompasses a wide range of theories and methods which may be viewed as offering tools for solving different problems in chemistry. There are three basic methods describing computational chemistry, namely quantum mechanics, molecular mechanics and molecular dynamics. The former focuses on electronic structure methods and the latter on classical laws of physics. In this chapter, mention is made of both computational and theoretical tools used in this study. This chapter begins with a description of the computational chemistry concepts that are central to all three delineated computational basic approaches, and these include; Schrodinger and Born-Oppenheimer equations and Potential energy surface (PES), followed by explanation of the both the computational and theoretical tools used in this our studies.

3.2. The Schrodinger equation

In 1926, an Austrian physicist by the name of Erwin Schrodinger, first proposed the time-independent Schrodinger equation, which provides a description of quantum system evolution [1]. The Schrodinger equation in its simplest form is as follows:

$$H = T + V \qquad \text{Eq. 3.1}$$

Where H is the Hamiltonian operator (sum of the kinetic energy), T is the potential energy, V is the operator [2, 3].

If particles are measured in terms of point masses based upon the assumption of assumed neglecting relativistic effects, then the Hamiltonian for a particular molecule will be composed of all kinetic and potential energy operators for all its electrons and nuclei. The Hamiltonian equation is as follows:

$$H = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{\hbar^2}{2}\sum_A \frac{1}{M_A}\nabla_A^2 - \sum_A\sum_i \frac{Z_A e^2}{r_{Ai}} + \sum_i\sum_{j>i}\frac{e^2}{r_{ij}} + \sum_A\sum_{B>A}\frac{Z_A Z_B e^2}{R_{AB}}$$ Eq. 3.2

As the Schrodinger equation is composed of several sets of equations, the disadvantage is that it cannot be applied to solve molecular system problems instead the Born-Oppenheimer approximation [1, 3, 4] serves the purpose.

3.3. Born-Oppenheimer approximation

An approximation, conducted by Born and Oppenheimer in 1927, explained that nuclear coordinates are the parameters that define molecular geometry [5] as well as highlighted the ratio between energy and nuclear coordinates, as described in PES. The Born-Oppenheimer approximation expounds that the Schrodinger equation for a specific molecular system may be split into two individual equations namely, electronic and nuclear equations. As nuclei are fixed, so the nuclear kinetic energy operator is deleted, and thus statically distributed electron contributions with a molecule can be solved.

$$T^{elec} = \left[-\frac{\hbar^2}{8\pi^2 m}\sum_i^{electrons}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\right]$$ Eq. 3.3.

Below is the Schrodinger equation for fixed nuclei electrons, which can be presented as:

$$H^{elec}\varphi^{elec}(r,R) = E^{eff}(R)\varphi^{elec}(r,R)$$ Eq. 3.4

3.4. Potential energy surface

The potential energy surface (PES) is a mathematical function that relates the energy of a molecular system in terms of its geometry [6]. This PES ratio provides a deeper visual insight into understanding the relationship between potential energy and its geometry of a molecule. It also affords researchers a chance to understand the workings of computational chemistry programs in order to isolate and characterize specific models of interest [6-8].
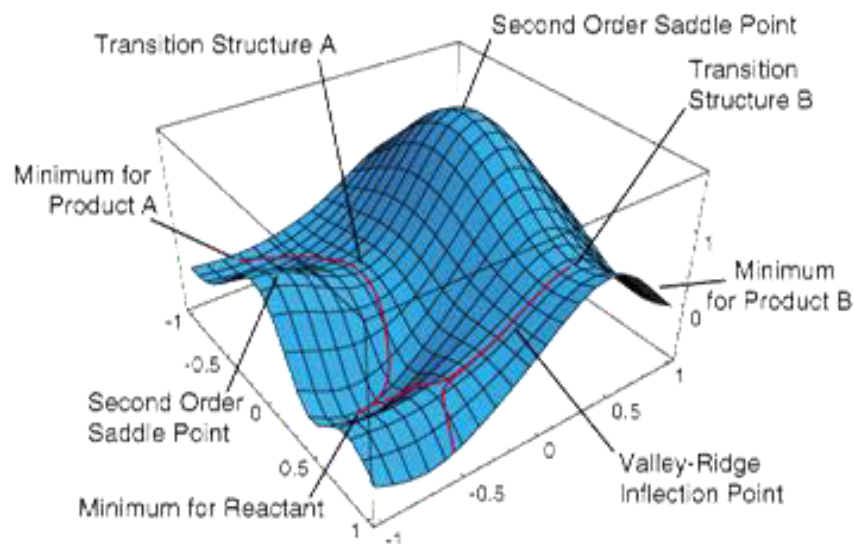
Figure 3.1. A model representation of a two-dimensional potential energy surface [9]

PES is represented as a graphical plot expressed as the ratio between molecular energy verses molecular geometry (**Figure 3.1.**). Molecular potential energy is directly proportional to its molecular geometric conformations. High energy regions depicts high-energy molecular conformational changes with unfavorable electronic interactions, whereas, regions of low potential energy expounds low energy molecular conformations with favorable electronic interactions [7, 8 10].

3.5. Quantum mechanics

Quantum is derived from the Latin word *quantus* which means "how much" and the first reported account in 1990, is believed to have been by Max Planck, who used quantum to demonstrate the absorption and emission spectra of energy as a measure of constrained particle quantities [1, 2, 6, 11]. However, the word mechanics denotes how an object would behave when a force (i.e. gravitational force) is applied. Molecules are composed of neutrons, electrons and nuclei, and therefore quantum chemistry studies electron motion when nuclear charges exert different electromagnetic forces in molecules [1, 2, 6, 11]. Thorough knowledge of quantum chemistry and in particular understanding of the Schrodinger equation is needed in order to gain a deeper insight into molecular electron behavior and as well as molecular structures and

reaction. Because in quantum chemistry, system properties are expressed as a wave function, which is solved by the Schrodinger equation[1, 2, 6, 11].

3.6. Molecular mechanics

Molecular mechanics (MM) is based on classical physics and the fundamental basis of modern physics is quantum mechanics. MM is commonly employed for large systems in order to predict relative potential energies of molecules in terms of molecular conformational changes or atom rearrangements [3, 4, 6]. MM may also be used to predict the equilibrated geometrical and transition states of molecules and relative energies of various conformers [12]. Although many of the details of mechanical, chemical and biological interactions in cells are currently unclear, MM can provide understanding of the mechanical deformation of proteins and nucleic acids and provide key insights for understanding the changes in cellular structure, response and function under force [13]. This may serve as a useful tool for diagnostic and therapeutic interventions in many diseases. MM may be simply described as a ball-and-spring model of atoms and molecules that have classical forces exerted between each other. MM calculations, often referred to as force field calculations, reveals the qualitative nature of how molecules interacts with one another [14].

3.6.1. Force field

A force field is a mathematical function with a delineated set parameters, which describes the relationship of molecular energy systems to specific particle coordinates. Some of the commonly used biomolecular force fields include: AMBER [15], CHARM [16], NAMD [17], GROMOS [18], and OPLS-AA [19]. All force field parameters are obtained from either experimental data sources, *ab initio* or semi-empirical QM [20]. There is an assumption that force field functions are solely dependent on atomic orientations and have been used as representatives for providing description of PES for different types of molecular systems with varying degrees of freedom. This force field function must provide a deeper understanding into the particular forces acting within the molecular system. Among these, is the valence force field function, applied for organic molecules, and may be used to calculate the energy of these molecules governed by the following set of parameters: internal coordinates, bond lengths, bond angles, dihedral angels and Cartesian co-ordinates of atoms [19], and presented as:

$$E_{tot} = E_{str} + E_{bend} + E_{tor} + E_{vdw} + E_{elec} \qquad \text{Eq. 3.5}$$

In these studies, the AMBER Force Field [21] was used, the standard AMBER Force Field was utilized for proteins, and however, the General AMBER Force Field (GAFF) was used to parameterize ligands.

3.7. Molecular dynamics

Molecular dynamics (MD) is a computational chemistry method used for molecular simulations of atoms in a modeled system with a given mass and charge. It also integrates energy equations of motion which are numerically solved to trace the time progression of a particular system under the assessment of kinetic and thermodynamic properties [22, 23]. The stimulating system has MM or QM methods applied. MD simply applies Newton's equations of motion to atoms on an energy surface for generating atomic trajectories of a complex molecular system;

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \qquad \text{Eq. 3.6}$$

Where $r_i(t) = (x_i(t), y_i(t), z_i(t))$ – is the particle position vector (i), t – time, m_i – mass of particle, therefore F_i - force acting on (i) particle at t and mi.

For assimilation of the above-mentioned second order linear differential equations, specification of instant particle forces, positions and velocities are required [20]. The Verlet algorithm is applied to solve the equations of motion [24]. The time progression of a molecular system and MD trajectories are outlined by position and velocity vectors, which are propagated using numerical integrators at specific time frames. This $r_i(t)$ describes a particle orientation space that alternates with time as compared to $V_i(t)$ which explains a system's thermodynamics (i.e. kinetic energy and temperature). The functional properties of a stimulating system that may be affected by a dynamic cascade of events may be directly identified at the atomic level [25].

3.7.1. Binding free energy

Binding free energy calculations are important in ligand-protein systems interactions and formation, and also in other aspects of computational drug design discovery including protein-

protein system behaviors and determining protein structures [26, 27, 28, 29-31]. Computer-based free binding energies of macromolecules and other varying molecular systems are calculated using either the Molecular Mechanics/Poisson-Boltzmann Surface Area (MMPB-SA) or Molecular Mechanics/Generalized Born Surface Area (MMGB-SA) techniques [25, 32, 33], which combines MM with the continuum solvation of a molecular system at low computational cost [33]. In the current studies, MM-PBSA free energy calculations were applied for the diverse molecular systems encountered [34]. A detailed explanation on free binding energy calculations has been described herein (**see Chapters 7-9**).

3.8. Hybrid QM/MM method

To reduce negative implications associated with using QM and MM approaches, therefore computational chemists have established combinational QM/MM algorithms, which merge QM descriptors with the low cost computing rate of MM [35]. As shown in **Figure 3.2**, is a model representation of an hybrid QM/MM approach, which is separated into two domains namely, the reactive domain defined by QM, whereas, the non-reactive domain treated with MM [36-39]. While QM calculations are very accurate at describing the electronic structure of molecular systems, it is highly computationally expensive and time consuming [40]. Combined QM/MM procedures intend to accurately describe quantum mechanics with the low cost applied to molecular mechanics computationally (*33*).
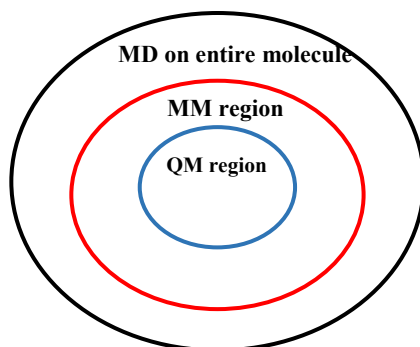
Figure 3.2. A schematic representation of a hybrid QM/MM/MD model

Fluctuating electronic rates in molecular systems and ligand-receptor interactions in metal-containing systems cannot be explained by classical MM approaches. Hybrid QM/MM may not relevant in all structure-based drug design processes, however, other crucial molecular systems

may not be well described by other computational methods, thus hybrid QM/MM is an imperative computational tool in drug design.

3.9. Principle component analysis

Classical MD analysis can only provide limited insight into the dynamic landscape frameworks of highly large and complex biomolecular systems. Therefore, other diverse computational methods have been established that can deal with the large number of explicit degrees of freedom [41-44]. One such widely used approach is principle component analysis (PCA), which explores the structural fluctuations experienced within biological systems.

PCA is a mathematical statistical technique based on covariance analysis that can transform the original space of correlated variables into reduced space of independent variables and is often used to highlight the large correlated motions in biological systems [45, 46]. PCA reduces the dimensionality of data from an MD simulation trajectory in order to extract dominant nodes accountable for pronounced molecule or protein motions that corresponds to correlated vibrational nodes [47-52]. A particular protein movement is denoted by an eigenvector with a specific corresponding energy contribution eigenvalue component. If a specific MD trajectory is projected onto a particular eigenvector, it describes the time-dependent movements that components undergo during a specific vibrational mode [48].

Herein, PCA was implemented for detection of the overall motion of the enzyme backbone to elucidate the difference between wild type protein states compared to mutant/resistant variant states (**see Chapter 7**) [53]. It was also used to reduce the dimensionality of data from an MD simulation to extract dominant modes in Nef bound and apo protein systems (**see Chapters 8 and 9**).

3.10. Residue Interaction Network

It has been reported that either resistant mutant forms or substitution in a target protein may affect the binding affinity of drugs to specific targets through modification of their residue interaction networks (RINs) [54-57]. Few studies have demonstrated that analyzing and visualizing

the RIN of proteins provided deeper insight on vital biological interactions in highly complex molecular systems [58, 59].

The RINalyzer software, a Java plugin for Cytoscape, is used for analysis and visualization of RINs in different proteins and biological systems. It also provides comprehensive insight into analyzing structural and functional properties or connections in different systems. In general, a RIN is created from a 3D protein structure as located in Protein Data Bank (PDB) files. RINalyzer permits one to concurrently visualize and explore the interactive 2D RIN in Cytoscape with a corresponding 3D representation in the molecular viewer UCSF Chimera program. The basic layout of an interaction network displays network nodes that represents various amino acid residues, whereas edges denotes respective molecular interactions between these amino acids.

3.11. Molecular modelling tools used in this study

3.11.1. Homology modeling

The availability of an actual X-ray crystal structure of a particular target of interest is critical in drug design discovery. Most targets can be readily downloaded from numerous protein websites, for example, the PDB and Drug Bank etc. However, previously if a specific target was not available, then scientists resorted to alternative powerful biophysical techniques namely, X-ray crystallography, NMR spectroscopy and electron microscopy for protein structure determination [60]. Despite the fact that using X-ray crystallography was the best approach to determine nearly accurate and high-resolution protein structures, but over the years, numerous implications have drastically affected the implementation of this methodology; such as increased trial-and-error and experimental time rates needed, as well as the lack of X-ray crystal structures of some biological important targets.

Therefore, homology modeling has been implemented, which predicts three dimensional (3D) protein structures that have little or no high resolution structural information by using closely related X-ray protein crystal structures with known structural information as homologues and protein sequencing data [61]. Protein resolution is defined as an insight into the degree of

diffraction patterns and electron density calculations. So protein structures of lower resolution (i.e. >3 Å) only display basic atomic framework details as compared to higher resolution structures (i.e. ~1 Å), where one can probe deeper into the atomic electron density map and are well-arranged. Thus, it is suggested that proteins with well-defined crystallographic structures (i.e. 1-3 Å) should be selected as homologues.

Herein, we introduced the concept of homology modeling in one of our studies (**Chapter 6**). As illustrated in **Figure 3.3** below, the actual human CCR5 enzyme was homology modeled using crystal structure CXCR4 (PDB: **3ODU**) as a structural homologue and the human CCR5 protein sequence retrieved from the UniProt database (Uniprot ID: **P51681**).
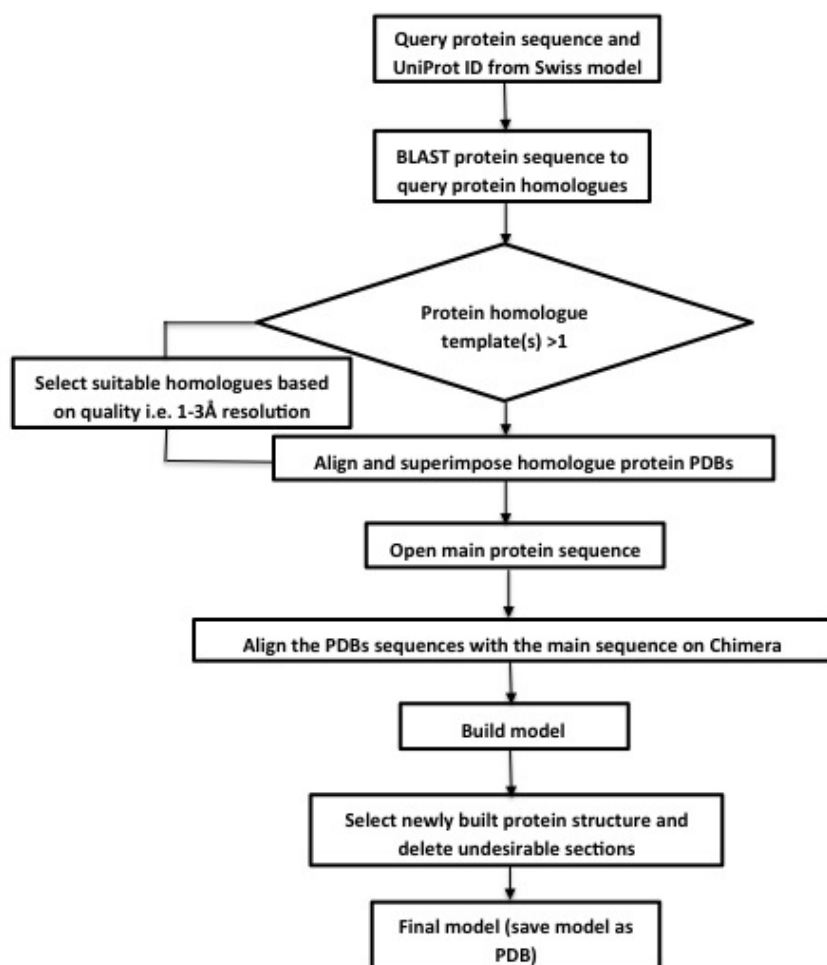
Figure 3.3. Schematic illustration of the protocol followed for developing the CCR5 homology model

3.12. References

1. Schroedinger, E. (1926) An undulatory theory of the mechanics of atoms and molecules, *Physical Review 28*, 1049 - 1070.

2. Jensen, F. (1999) Introduction to computational chemistry, *Wiley and Sons 2nd Ed.*

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

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

5. Born, M., and Oppenheimer, R. (1927) Zur Quantentheorie der Molekeln *Annals of Physics 84*, 457-484.

6. Atkins, P., de Paula, J. (2006) Atkin's Physical Chemistry, *8 Ed.*

7. Elsawy, K. M., Hodgson, M. K., and Caves, L. S. D. (2005) The physical determinants of the DNA conformational landscape: An analysis of the potential energy surface of single-strand dinucleotides in the conformational space of duplex DNA., *Nucleic Acids Reserach 33*, 5749-5762.

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

9. Grobler, J. A., Stillmock, K., Hu, B. H., Witmer, M., Felock, P., Espeseth, A. S., Wolfe, A., Egbertson, M., Bourgeois, M., Melamed, J., Wai, J. S., Young, S., Vacca, J., and azuda, D. J. (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: Implications or metal binding in the active site of phosphotransferase enzymes, *Proceedings of the National Academy of Sciences U. S. A. 99*, 6661-6666.

10. Atkins, P., and Friedman, R. (2005) Molecular quantum mechanics, *Oxford University Press Inc. 4th Ed.*

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

12. Lewars, E. G. (2011) *Computational Chemistry*, Springer Netherlands.

13. Bao, G. S., S. (2003) Cell and molecular mechanics of biological materials, *Nature 2*, 715-725.

14. Senn, H. M., Thiel, Walter.,. (2009) QM/MM methods for biomolecular systems, *Angewandte Chemie International Edition 48*, 1198-1229.

15. 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.

16. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program, *Journal of Computational Chemsitry 30*, 1545-1614.

17. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *Journal of Computational Chemistry 26*, 1781-1802.

18. Christen, M., Hunenberger, P. H., Bakowies, D., Baron, R., Burgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Krautler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and Van Gunsteren, W. F. (2005) The GROMOS software for biomolecular simulation: GROMOS05, *Journal of Computational Chemistry 26*, 1719-1751.

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.

20. González, M. A. (2011) Force fields and molecular dynamics simulations, *EDP Sciences 12*, 169–200.

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-programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free-energy calculations to simulate the structural and energertic properties of a molecule, *Computer Physics Communications 91*, 1-41.

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

23. Karplus, M. (2003) Molecular dynamics of biological macromolecules: A brief history and perspective, *Biopolymers 68*, 350-358.

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

25.   Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions, *Annual Review of Biophysics and Biomolecular Structure 30*, 211-243.

26.   Kamerlin, S. C. L., Haranczyk, M., and Warshel, A. (2009) Progress in Ab Initio QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pK(a), Redox Reactions and Solvation Free Energies, *Journal of Physical Chemistry B 113*, 1253-1272.

27.   Kollman, P. (1993) Free-energy calculations: Applications to chemical and biochemical phenomena, *Chemical Reviews 93*, 2395-2417.

28.   Homeyer, N., and Gohlke, H. (2012) Free energy calculations by the molecular mechanics poisson−boltzmann surface area method, *Molecular Informatics 31*, 114−122.

29.   Gohlke, H., and Case, D. A. (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf, *Journal of Computational Chemistry 25*, 238-250.

30.   Hou, T. J., Chen, K., McLaughlin, W. A., Lu, B. Z., and Wang, W. (2006) Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain, *PLoS Computational Biology 2*, 46-55.

31.   Wang, W., and Kollman, P. A. (2000) Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model, *Journal of Molecular Biology 303*, 567-582.

32.   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.

33.   Wang, J. M., Hou, T. J., and Xu, X. J. (2006) Recent advances in free energy calculations with a combination of molecular mechanics and continuum models, *Current Computer-Aided Drug Design 2*, 287-306.

34. Kar, P. K., V. (2012) Origin of decrease in potency of darunavir and two related antiviral inhibitors against HIV-2 compared to HIV-1 protease, *Journal of Physical Chemistry B 116*, 2605−2614.

35. Lin, H., and Truhlar, D. G. (2007) QM/MM: what have we learned, where are we, and where do we go from here?, *Theoretical Chemistry Accounts 117*, 185-199.

36. Garciavela, A., and Gerber, R. B. (1993) Hybrid quantum/semiclassical wave packet method for molecular dynamics: Application to photolysis of Ar-HCl *Journal of Chemical Physics 98*.

37. Sauer, J., and Sierka, M. (2000) Combining quantum mechanics and interatomic potential functions in ab initio studies of extended systems, *Journal of Computational Chemistry 21*, 1470-1493.

38. Monard, G., Prat-Resina, X., Gonzalez-Lafont, A., and Lluch, J. M. (2003) Determination of enzymatic reaction pathways using QM/MM methods, *International Journal of Quantum Chemistry 93*, 229-244.

39. Menikarachchi, L. C., and Gascon, J. A. (2010) QM/MM approaches in medicinal chemistry research, *Current Topics in Medicinal Chemistry 10*, 46-54.

40. Honarparvar, B., Kruger, H. G., Maguire, G. E. M., Govender, T., and Soliman, M. E. S. (2012) Prospects of computer modeling combined with spectroscopic techniques for the analysis of enzyme-inhibitor interactions: Review of an integrated approach for structure-based drug design *Chemical Review, accepted, 2013*.

41. Cheng, X. L., Cui, G. L., Hornak, V., and Sinnnerling, C. (2005) Modified replica exchange simulation methods for local structure refinement, *Journal of Physical Chemistry B 109*, 8220-8230.

42. Affentranger, R., Tavernelli, I., and Di Iorio, E. E. (2006) A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling, *Journal of Chemical Theory and Computation 2*, 217-228.

43. Okur, A., Wickstrom, L., Layten, M., Geney, R., Song, K., Hornak, V., and Simmerling, C. (2006) Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model, *Journal of Chemical Theory and Computation 2*, 420-433.

44.   Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water, *Proceedings of the National Academy of Sciences of the United States of America 102*, 13749-13754.

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

46.   Thomas, J. R., Gedeon, P. C., Grant, B. J., and Madura, J. D. (2012) LeuT Conformational Sampling Utilizing Accelerated Molecular Dynamics and Principal Component Analysis, *Biophysical Journal 103*, L01-L03.

47.   Amadei, A., Linssen, A. B., de Groot, B. L., van Aalten, D. M., and Berendsen, H. J. (1996) An efficient method for sampling the essential subspace of proteins, *Journal of Biomolecular Structure and Dynamics 13*, 615-625.

48.   van Aalten, D. M., Findlay, J. B., Amadei, A., and Berendsen, H. J. (1995) Essential dynamics of the cellular retinol-binding protein– evidence for ligand-induced conformational changes, *Protein Engineering, Design and Selection 8*, 1129-1135.

49.   Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins, *Proteins: Structure, Function, and Bioinformatics 17*, 412-425.

50.   Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms, *Journal of Molecular Biology 285*, 1711-1733.

51.   Case, D. A. (1994) Normal-mode analysis of protein dynamics, *Current Opinion in Structural Biology 4*, 285-290.

52.   Brooks, B., and Karplus, M. (1985) Normal-modes for specific motions of macromolecules - Application to the hinge-bending mode of lysozyme, *Proceedings of the National Academy of Sciences of the United States of America 82*, 4995-4999.

53.   Laine, E., de Beauchene, I. C., Perahia, D., Auclair, C., and Tchertanov, L. (2011) Mutation D816V Alters the Internal Structure and Dynamics of c-KIT Receptor Cytoplasmic Region: Implications for Dimerization and Activation Mechanisms, *Plos Computational Biology 7*.

54.   del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins, *Molecular Systems Biology 2*.

55.   Welsch, C., Schweizer, S., Shimakami, T., Domingues, F. S., Kim, S., Lemon, S. M., and Antes, I. (2012) Ketoamide Resistance and Hepatitis C Virus Fitness in Val55 Variants of the NS3 Serine Protease, *Antimicrobial Agents and Chemotherapy 56*, 1907-1915.

56.   Welsch, C., Domingues, F. S., Susser, S., Antes, I., Hartmann, C., Mayr, G., Schlicker, A., Sarrazin, C., Albrecht, M., Zeuzem, S., and Lengauer, T. (2008) Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus, *Genome Biology 9*.

57.   Xue, W. W., Jin, X. J., Ning, L. L., Wang, M. X., Liu, H. X., and Yao, X. J. (2013) Exploring the Molecular Mechanism of Cross-Resistance to HIV-1 Integrase Strand Transfer Inhibitors by Molecular Dynamics Simulation and Residue Interaction Network Analysis, *Journal of Chemical Information and Modeling 53*, 210-222.

58.   Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011) Analyzing and visualizing residue networks of protein structures, *Trends in Biochemical Sciences 36*, 179-182.

59.   Doncheva, N. T., Assenov, Y., Domingues, F. S., and Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures, *Nature Protocols 7*, 670-685.

60.   Nollmann, M., Stark, W. M., and Byron, O. (2005) A global multi-technique approach to study low-resolution solution structures, *Journal of Applied Crystallography 38*, 874-887.

61.   Pitman, M. R., and Menz, R. I. (2006) Methods for protein homology modelling, *Bioinformatics 6*, 37- 59.

CHAPTER 4

Dual Acting HIV Inhibitors: Integrated Rational *in-silico* Design Strategy.

Suri Moonsamy[a] and Mahmoud E. S. Soliman[a*]

[a]School of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

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

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

4.1. Abstract

In-silico pharmacophore and structure-based drug design guided by binding mode analysis and molecular dynamics simulations is an effective approach for identifying novel potent and selective inhibitors. Herein, we demonstrated a unique strategy for developing dual acting inhibitors against HIV-1 protease (PR) and reverse transcriptase (RT). Via an integrated computational protocol relies on "loop docking" and molecular dynamics simulations, the designed targets exhibited binding affinities comparable to, and in some cases better than, known active reference drugs.

Keywords: Dual acting HIV inhibitors, protease inhibitors, reverse transcriptase inhibitors, computer-aided drug design

4.2. Introduction

Three main HIV-1 enzymes have been exploited as the core of chemotherapy for AIDS treatment. These enzymes are HIV-1 protease (PR), reverse transcriptase (RT) and integrase (IN) [1, 2]. However, designing inhibitors to target each single enzyme has proved a successful strategy for treatment, the therapeutic effect of these inhibitors is drastically hampered by many implications. Nucleoside RT inhibitors (NRTIs) exhibited high intrinsic toxicity [3], whereas the less toxic protease inhibitors (PIs), non-nucleoside RT inhibitors (NNRTIs) and integrase inhibitors (INIs) are severely compromised by the rapid emergence of resistant viral strains [4-6].

A common tactic to overcome HIV disease resistant to a single therapy is the use of multidrug therapy. Highly Active Antiretroviral Therapy (HAART) [7] combines NRTIs with NNRTIs or PIs and effectively reduces HIV viral load which in turn considerably improves the life quality of AIDS patients [8]. Even though, HARRT proved successful, to ensure the efficacy of this primary treatment for AIDS, complete adherence - which is very difficult to attain because of complicated dosing and intolerable acute and chronic toxicities - is essential. Failure to achieve adherence typically results in viral rebound and, even worse, multidrug resistance.

To this end, we envisage a dual acting anti-HIV therapy using a single drug that concomitantly inhibit two viral enzymes could result in lower toxicity, simplified dosing, and improved patient adherence, therefore reducing the possibility of drug resistance.

The concept of designing of multifunctional ligands has been proposed before [9]. Multifunctional ligands were classified into three categories on the basis of how the scaffolds were combined [9]: (a) conjugate, if scaffold are linked by a linker, (b) fused, if they directly coupled with no linker, and (c) merged, if the scaffolds are integrated into a single structure sharing common feature. Another classification based on the targeted protein binding sites was also introduced. Targeted protein binding sites could be: (a') adjacent within a single protein structure [10, 11], (b') located in different proteins but recognize similar endogenous ligands [12, 13], and (c') located in different proteins but recognize unrelated ligands [14-16]. In this report, we propose potential inhibitors based on the above-mentioned (c and c') classifications.

In this study, we aimed to design novel inhibitors with potential dual activity against HIV PR and RT enzymes. Our design is initially based on the pharmacophoric features conserved in most potent FDA-approved PR and NRT inhibitors (**Figure 4.1.**). Then, the pharmacophore-based created scaffolds were subjected to structure-based screening against the respective enzyme to estimate their binding affinities.
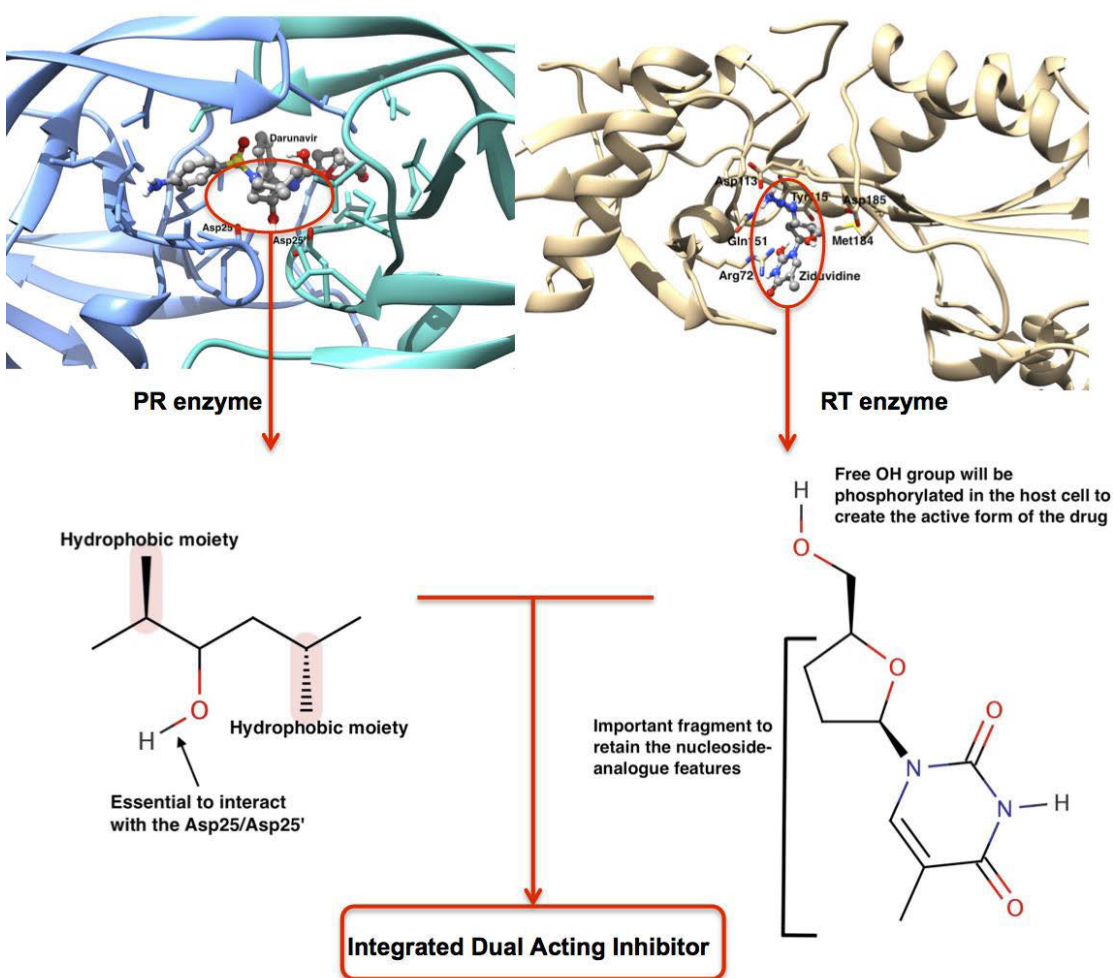
Figure 4.1. The design of dual acting inhibitors is based on the pharmacophoric features of known inhibitors for each enzyme target and the binding theme of these pharmacophoric groups with the respective enzyme active site. Darunavir-PR complex (PDB code: 3QOZ) and Zidovudine-RT complex (PDB code: 3KLF) were used in this study to assign the pharmacophoric moieties.

In the current work, the strategy to design potential dual acting inhibitors is assisted by an integrated in-silico approach that exploits the structural features of both the inhibitor and the enzyme active site. With no doubt that the most appropriate way to verify the activity of *in-silico* designed target is the experimental *in-vitro* and/or *in-vivo* testing, however, the time and cost factors should be considered carefully before testing anticipated inhibitors. Validated computational tools could serve as useful tools to save time and effort in the drug design process. In this report, we introduce a "tough" *in-silico* approach that relies on the use of docking calculations followed by molecular dynamics simulations to accurately estimate the binding affinity as well as the stability of the inhibitor-enzyme complexes. As a matter of fact, we believe that docking calculations could be entirely misleading and even using the same software, in many case, may result in different predictions. In order to ensure reliable docking results we embarked on a novel approach which we name "loop-docking" to enhance the docking calculations predications (**see Methods sections**). Furthermore, the docked structures were then subjected to molecular dynamics simulations to ensure that these docked structure are stable enough over a reasonable time scale. Docked structures that fail to exhibit stability during MD simulations are subjected to further docking or rejected. The "loop docking"/MD strategy have been tested and proved successful on 30 X-ray crystal structures of various ligand-protein complexes (A manuscript that fully describe this approach has been submitted and under review). Herewith, we report the main concept and structural scaffolds of proposed dual acting PR/RT HIV inhibitors. Screening of a wider set of substituents for further lead optimization is still running. Synthesis of these compounds is quite feasible and most starting structures are commercially available. Since the synthesis and testing of the proposed inhibitors is still on-going, detailed information about the synthetic procedure, structural analysis and biological testing will be revealed in a forthcoming publication. Previously, we have released the X-ray crystal structure of the HIV PR subtype C [17] as well as several reports on the synthesis, NMR and biological testing of a series of anti-HIV PR inhibitors [18-22]. Also, we demonstrated an extensive study on the relative binding affinities of nine anti-HIV FDA-approved drugs against different PRs subtypes [17].

The unique strategy adopted in this report could serve as a powerful tool in the process of drug design and development, not only against HIV, but for a wide range of biological targets as well.

4.3. Methods: in-silico design and validation

4.3.1. Constructing of the proposed structures and conformational ensembles

Proposed structures were constructed and energy minimized using Avogadros software [23]. To ensure through conformational sampling of the ligand candidates, low-energy conformations were generated for each molecule using Omega v1.8b [24]. For each compound, conformers generated within a range of 3.0 kcal/mol from the lowest energy conformer were investigated.

4.3.2. Protein systems

Protein structures were obtained from the protein data bank (http://www.rcsb.org/). The PDB codes 3QOZ and 3KLF were used for PR and RT, respectively. Procedure for enzyme model preparation for docking and MD simulations is explained in details elsewhere [18-20].

4.3.3. Docking calculations: "loop docking"

The idea behind our in-house "loop docking" approach is based on the fact that consecutive docking runs could remarkably improve the docking energy and orientation. From our experience with docking, we realized that in many cases the best-docked structure could be a docking artifact and does not represent the best docking orientation. Therefore, we always opt to re-run the docking calculation using the best-docked structure from initial docking as a starting structure for a second docking run and so on. We used a few scripts to allow this process to be automated. This automated "loop docking" will continue until a threshold value is reached. The threshold value is the difference between the docking binding energy of the last run and the preceding one. A threshold value of 0.05 was found to be appropriate. When this imposed value reached, the docking stops and the best-docked structure is selected. Autodock software [25] was used for docking calculations. Details on docking set up and parameters can be found in our previous reports [18-20].

4.3.4. Molecular dynamics (MD) simulations

As before [17], MD simulations were performed for all inhibitor-enzyme complexes (8 systems in total) using Amber software [26].

4.4. Results and Discussion

4.4.1. Design plan: pharmacophore and structure-based design

Our design began with the chemical knowledge of the structural features of PR and RT inhibitors as well as the anatomy of the binding sites of both enzymes. First, we carefully studied the binding mode for all up-to-date resolved HIV PR and RT X-ray crystal structures of ligand-enzyme complexes in order to determine the common and most conserved pharmacophoric groups that critically involved in the ligand binding. However, finding a specific feasible way to help us easily analyze these ligand-enzyme binding themes was not an easy task to accomplish. After many trials, we came up with an efficient and viable procedure. For each enzyme class, PR and RT, all X-ray crystal structures were aligned using Chimera software [27]. This enabled us to easily focus only on the bound ligands to determine the common structural and binding components. Then, Ligplot software [28] was used to create a visual map of the most critical binding interactions between the ligands and enzyme active site. Information gained from this extensive structural and binding analysis allowed us to define the most important fragments for each class of inhibitors, PRIs and NRTIs (see **Figure 4.1**). For PRIs, a characteristic structural feature is that hydrogen bond donors in the core of the ligand is important to form a hydrogen bond with the active site Asp25/Asp25' residues. It has been shown that the interaction between the two catalytic aspartates and the hydroxyl group is worth more than 4 kcal/mol [29]. Another feature is that the PR enzyme S1/S1'subsites are mostly hydrophobic and ligand with hydrophobic groups that can fill in theses subsites have showed significant enzyme inhibitory activity [30]. Some studies have also reported that inhibitors that interact with the flap residues could be potential inhibitors [31].

The same previous procedure has been repeated for the X-crystal structures of RT-NRTIs complexes in order to understand the binding theme of these inhibitors and obtain information on the most critical pharmacophoric moieties. Unlike PR inhibitors, it turned out that the binding behavior of NRTIs slightly varies from one crystal structure to another, however, seven amino acid residues seemed to be mainly involved in the binding with the inhibitors. These amino acids are Arg72, Gln151, Met184, Ala114, Asp113, Tyr115 and Phe116.

Based on all the above-mentioned compiled information, we proposed main four structural scaffolds (see **Table 4.1**) that we believe it could exhibit dual acting HIV PR/RT inhibitory

activity. Lead optimization and expansion phases, in which a larger set of compounds will be evaluated, are on-going. As mentioned above, synthesis of these proposed targets is quite feasible and almost all starting structures are commercially available.

4.4.2. Docking and MD simulations

The docking binding free energies for the proposed structures (**1-4**) against PR and RT are shown in **Table 4.1**. Docking of reference drugs, Darunavir and Ziduvidine, was also performed for comparison (**Table 4.1**).

The docked complexes were then subjected to MD simulation for 5 ns to ensure the stability of the ligands in the enzyme active sites. Form our previous experience with docking, in many cases, we found that, even the best docked structure flies away from the enzyme active site within a few picoseconds of MD simulations. Therefore, we believe that docking calculations that are not validated by relatively a long MD run to ensure the stability of the system is not reliable. Interestingly, for all the proposed compounds-enzyme complexes, the average RMSD values were below 2.3 Å and the variation of the potential energies falls within 1000 kcal/mol and this is a good indication of the system stability. The PDB coordinates for all docked complexes (8 in total) are provided with the supplementary material.

Table 4.1. 2D structures, the calculated binding energy and chemical properties based on Lipinski's rule of five.

Compound	2D structure	Calculated binding energy (kcal/mol)	
		PR	RT
		DRV	AZT
		(-9.4)*	(-6.2)*
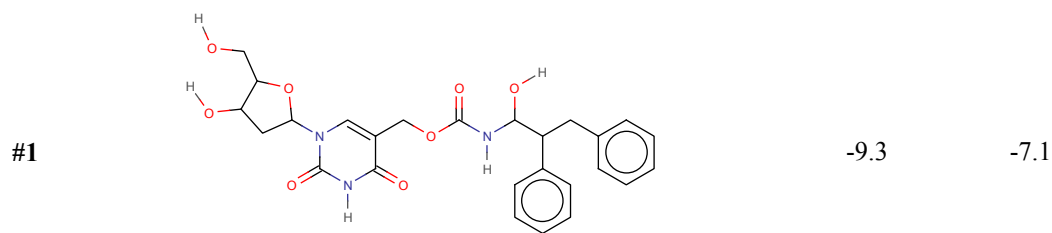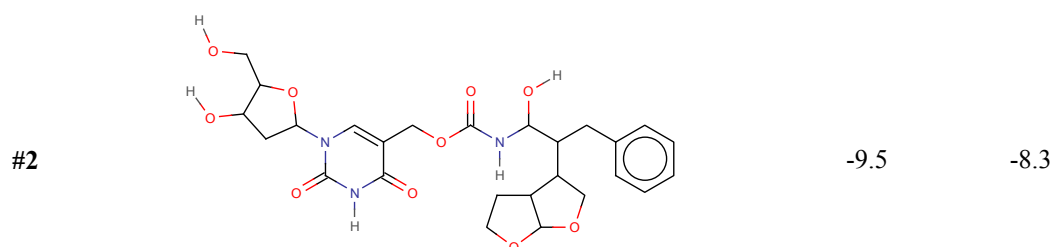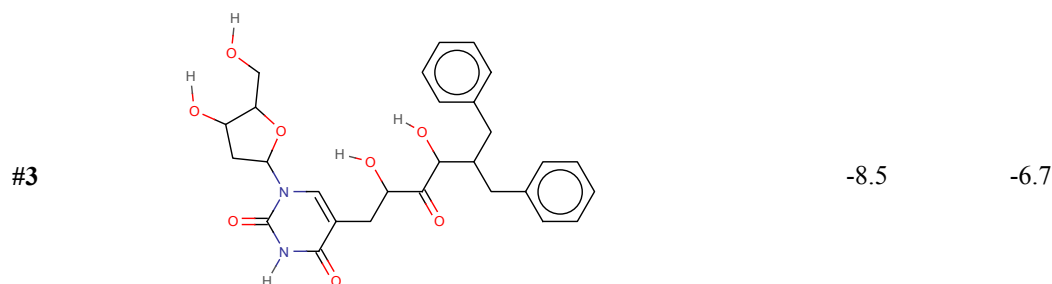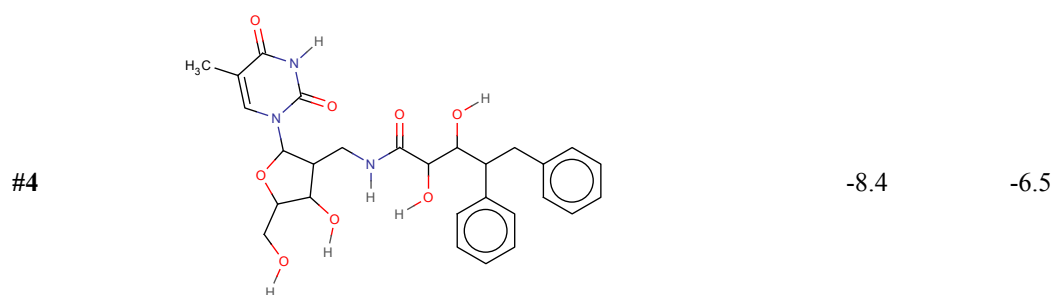
#1	**MWt (511.2), LogP(1.4), Ha (7), Hd(5)	-9.3	-7.1
#2	MWt (547.2), LogP(-0.13), Ha (9), Hd(5)	-9.5	-8.3
#3	MWt (524.5), LogP(1.7), Ha (8), Hd(5)	-8.5	-6.7
#4	MWt (539.5), LogP(0.50), Ha (9), Hd(6)	-8.4	-6.5

* Calculated binding energy for DRV and AZT against PR and RT enzymes, respectively.

** Calculated using MarvinSketch (http://www.chemaxon.com)

As evident from **Table 4.1.**, interestingly, structure **2** shows the best binding affinity towards PR and RT. This might due to the inclusion the bis-tetrahydrofuranyl moiety. The bis-tetrahydrofuranyl moiety is found in Darunavir structure and as we anticipated, inclusion of this moiety resulted in a structure (compound **2**) with high binding affinity comparable to Darunavir.

As also seen from **Table 4.1**, all the proposed structures show better binding affinity towards RT than Ziduvidine. This observation might be explained that inclusion of hydrophobic fragments in the structures increases hydrophobic interactions with the enzyme pocket residues and hence, improved the overall inhibitor binding.

4.4.3. Ligand-enzyme interactions

Post-dynamic complex structures were analyzed to ensure that the proposed structures fit well in the enzyme active pocket and, to some extent, exhibit similar interaction trends as the reference drugs with their respective enzyme targets. As evident from the hydrogen bonding and electrostatic interactions plots (**Figure 4.2.**), all proposed compounds showed, to a great extent, comparable interactions as in the crystal structures of the prototypes, DRV and AZT (all PDB coordinates are provided in the electronic supplementary material).
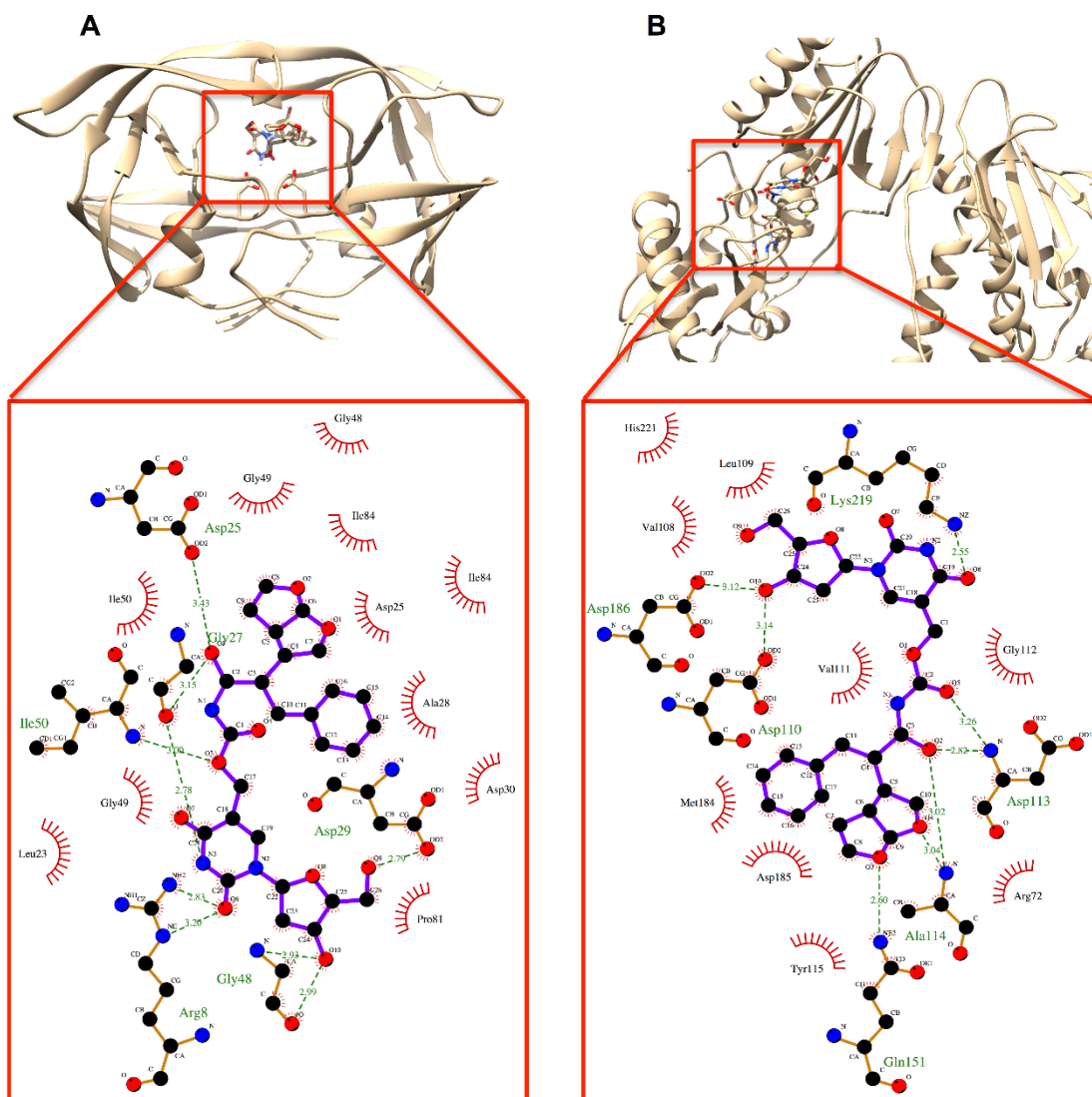
Figure 4.2. Compound 2 in complex with PR and RT, A and B, respectively, showing the hydrogen bonding and electrostatic interactions with the enzymes active site.

4.4.4. Per-residue interaction

For further lead optimization and expansion, it is useful to have a quantitative estimation of the contribution of each amino acid residues in the enzyme active site towards ligand binding. To accomplish this, we computed the per-residue interaction MolDock scoring function [32]. Insight obtained from per-residue interactions will allow us to tailor inhibitors that can strongly interact with those residues showed the most contribution towards the binding. Per-residue interactions for compounds **2** with PR and RT are shown in **Figure 4.3. A** and **B**.
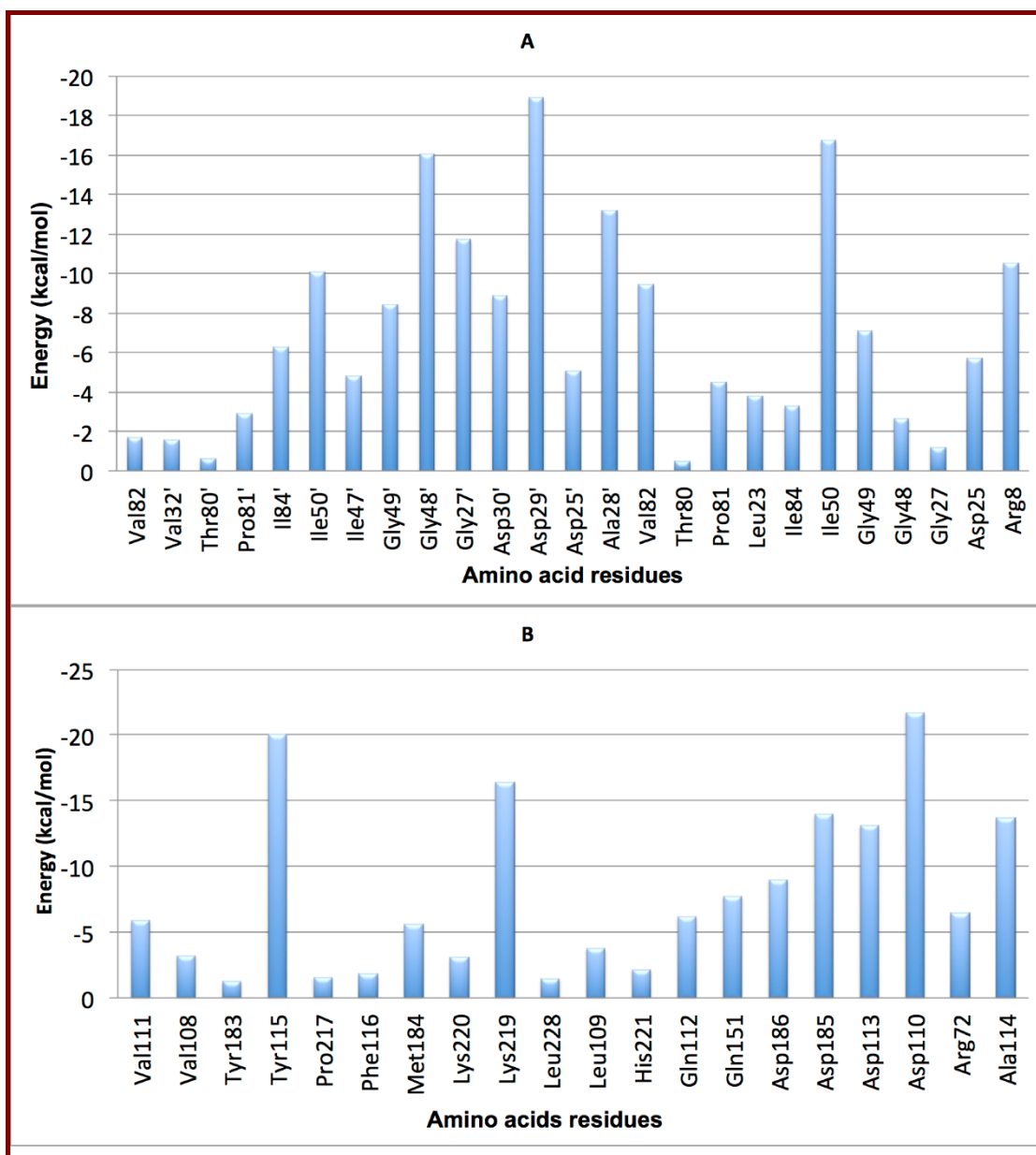
Figure 4.3. Per-residue interactions of compound 2 with PR and RT, A and B, respectively.

From the hydrogen bonding interaction plots and per-residue interaction calculations, inhibitors clearly exhibited strong interaction with PR enzyme via hydrogen bond with the active site residues Asp25/Asp25', a conserved interaction seen in all X-ray crystal structures of PIs-enzyme complexes. Interestingly, compound **2** showed a strong interaction with the flap residues, Ile50/Ile50'. Inhibitors that can interact with flap residues are believed to interfere with the flap opening and closure process, hence enzyme normal function [31]. With respect to RT,

proposed structures showed hydrogen bonding as well as hydrophobic interactions with active site pocket residues, most importantly with Tyr115, Asp110, Arg72 and Ly219 (**Figure 4.2. and 4.3.**). These interactions were also seen in the X-ray crystal structure of the AZT-RT complex.

As evident from the docking, MD simulations and per-residue calculations, the four proposed lead compounds are expected to be potential dual acting inhibitors and good starting point for further optimization and expansion.

The design strategy presented here could be implemented in the drug design and development process not only against HIV targets but also a wide range of biological systems

4.5. Conclusion

Based on the structural information of known HIV PR and RT inhibitors and the understanding of their binding mechanisms to the targets, novel leads were proposed as potential dual acting PR/RT inhibitors. Using hybrid computational approaches involving "loop docking" and molecular dynamics simulations and per-residue interactions calculations, the proposed structures were found to exhibit better binding affinities when compared to the prototype drugs. Compound 2 with bis-tetrahydrofuranyl moiety was found to be the most active compound against both PR and RT. The structures presented in this work are proposed as promising leads for developing potential drugs against HIV/AIDS therapy. Moreover, the *in-silico* strategy introduced in this study could be useful in the process of drug design and development.

4.6. References

1.   Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.

2.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets - Infectious Disorders 2*, 17-32.

3. Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses (vol 71, pg 322, 2006), *Antiviral Research 72*, 256-256.

4. Young, B., and Kuritzkes, D. R. (2002) Resistance to HIV-1 protease inhibitors., *Infectious Diseases and Therapy 25*, 257-282.

5. Bacheler, L. T. (1999) Resistance to non-nucleoside inhibitors of HIV-I reverse transcriptase, *Drug Resistance Update 2*, 56-67.

6. Mbisa, J. L., Martin, S. A., and Cane, P. A. (2011) Patterns of resistance development with integrase inhibitors in HIV, *Infection and drug resistance 4*, 65-76.

7. Yeni, P. (2006) Update on HAART in HIV, *Journal of Hepatology 44*, S100-S103.

8. Palella, F. J., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., Aschman, D. J., and Holmberg, S. D. (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection, *The New England Journal of Medicine 338*, 853-860.

9. Morphy, R., and Rankovic, Z. (2009) Designing Multiple Ligands - Medicinal Chemistry Strategies and Challenges, *Current Pharmaceutical Design 15*, 587-600.

10. Lewis, W. G., Green, L. G., Grynszpan, F., Radic, Z., Carlier, P. R., Taylor, P., Finn, M. G., and Sharpless, K. B. (2002) Click chemistry in situ: Acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks, *Angewandte Chemie International Edition 41*, 1053-+.

11. Munoz-Torrero, D., and Camps, P. (2006) Dimeric and hybrid anti-Alzheimer drug candidates, *Current Medicinal Chemistry 13*, 399-422.

12. Gemma, S., Gabellieri, E., Huleatt, P., Fattorusso, C., Borriello, M., Catalanotti, B., Butini, S., De Angelis, M., Novellino, E., Nacci, V., Belinskaya, T., Saxena, A., and Campiani, G. (2006) Discovery of huperzine A-tacrine hybrids as potent inhibitors of human cholinesterases targeting their midgorge recognition sites, *Journal of Medicinal Chemistry 49*, 3421-3425.

13. Yamamoto, M., Ikeda, S., Kondo, H., and Inoue, S. (2002) Design and synthesis of dual inhibitors for matrix metalloproteinase and cathepsin, *Bioorganic and Medicinal Chemistry Letters 12*, 375-378.

14. Kawanishi, Y., Ishihara, S., Tsushima, T., Seno, K., Miyagoshi, M., Hagishita, S., Ishikawa, M., Shima, N., Shimamura, M., and Ishihara, Y. (1996) Synthesis and pharmacological evaluation of highly potent dual histamine H-2 and gastrin receptor antagonists, *Bioorganic and Medicinal Chemistry Letters 6*, 1427-1430.

15. Ryckmans, T., Balancon, L., Berton, O., Genicot, C., Lamberty, Y., Lallemand, B., Pasau, P., Pirlot, N., Quere, L., and Talaga, P. (2002) First dual NK1 antagonists-serotonin reuptake inhibitors: Synthesis and SAR of a new class of potential antidepressants, *Bioorganic and Medicinal Chemistry Letters 12*, 261-264.

16. Groneberg, R. D., Burns, C. J., Morrissette, M. M., Ullrich, J. W., Morris, R. L., Darnbrough, S., Djuric, S. W., Condon, S. M., McGeehan, G. M., Labaudiniere, R., Neuenschwander, K., Scotese, A. C., and Kline, J. A. (1999) Dual inhibition of phosphodiesterase 4 and matrix metalloproteinases by an (arylsulfonyl)hydroxamic acid template, *Journal of Medicinal Chemistry 42*, 541-544.

17. S.M., A., 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, *D - 101262549 81*, 208-218.

18. Honarparvar, B., Makatini, M. M., Pawar, S. A., Petzold, K., Soliman, M. E., Arvidsson, P. I., Sayed, Y., Govender, T., Maguire, G. E., and Kruger, H. G. (2012) Pentacycloundecane-diol-Based HIV-1 Protease Inhibitors: Biological Screening, 2D NMR, and Molecular Simulation Studies, *ChemMedChem 7*, 1009-1019.

19. 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, *Bioorganic Chemistry 40*, 19-29.

20. Makatini, M., Petzold, K., Sriharsha, N. S., Ndlovu, N., Soliman, M. E. S., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Synthesis and Structural Studies of Pentacycloundecane-based HIV-1 PR Inhibitors: A Hybrid 2D-NMR and Docking/QM/MM/MD Approach, *European Journal of Medicinal Chemistry 46*, 3976-3985.

21. Makatini, M., Petzold, K., Sriharsha, S., Soliman, M. E. S., Honarparvar, B., Arvidson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H., and Govender, T. (2011) Pentacycloundecane-based Inhibitors of Wild-type C-South African HIV Protease *Bioorganic and Medicinal Chemistry Letters 15* 2274-2277.

22. Makatini, M. M., Petzold, K., Alves, C. N., Arvidsson, P. I., Honarparvar, B., Govender, P., Govender, T., Kruger, H. G., Sayed, Y., Jeronimolameira, Maguire, G. E., and Soliman, M. E. S. (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors, *Journal of Enzyme Inhibition and Medicinal Chemistry 28*, 78-88.

23. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An advanced semantic chemical editor, visualization, and analysis platform", *Journal of Cheminformatics 4*, 1-17.

24. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *ournal of Chemical Information and Modeling 50*, 572-584.

25. 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.

26. 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.

27. 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.

28. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Ligplot - A program to generate schematic diagrams of protein ligand interactions, *Protein Engineering 8*, 127-134.

29. Appelt, K. (1993) Crystal structures of HIV-1 protease-inhibitor complexes, *Perspectives in Drug Discovery and Design 1*, 23-48.

30. Brik, A., and Wong, C. (2003) HIV-1 protease: mechanism and drug discovery, *Organic and Biomolecular Chemistry 1*, 5-14.

31. Damm, K. L., Ung, P. M., Quintero, J. J., Gestwicki, J. E., and Carlson, H. A. (2008) A poke in the eye: inhibiting HIV-1 protease through its flap-recognition pocket, *89*, 643-652.

32. Thomsen, R., and Christensen, M. (2006) MolDock: a new technique for high-accuracy molecular docking, *Journal of Medicinal Chemistry 1*, 3315-3321.

CHAPTER 5

Computer-aided perspective for the design of flexible HIV non-nucleoside reverse transcriptase inhibitors (NNRTIs): *de-novo* design, virtual screening and molecular dynamics simulations

Suri Moonsamy[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. Soliman, email: soliman@ukzn.ac.za

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

5.1. Abstract

Flexible hydroquinone-based compounds were previously proposed as potential mutant-resistant NNRTIs inhibitors, however, experimental or computational evidences did not support this proposal. Herewith, using an integrated *in-silico* computational approach involving *de-novo* design, structure-based virtual screening (SBVS), molecular dynamics simulations and post-dynamic per-residue binding energy decomposition analysis, the binding affinity as well as the interaction landscape of novel flexible hydroquinone-based compounds were investigated to explore their activity as potential NNRTIs. The proposed leads were found to exhibit improved binding affinity when compared to FDA-approved NNRTIs, rilpvirine, nevirapine and efavirenz, however, the bioavailability profile of these compounds could hamper their uses as effective drugs. Results obtained from this extensive study could assist medicinal and biochemistry researchers with further experimental investigations.

5.2. Introduction

As the human immunodeficiency virus type 1 (HIV-1) has been proven to be the cause of the acquired immunodeficiency syndrome (AIDS) [1], this disease poses detrimental effects for world health as it is estimated that 34 million people live with HIV/AIDS worldwide [2, 3].

The reverse transcription (RT) enzyme is important for HIV-1 replication [4]. HIV-1 RT is a vital component in the virus life cycle, where this enzyme facilitates converting single-stranded RNA viral genome into double-stranded DNA [4-6]. As RT is considered as a critical component in the virus life cycle, this HIV protein has become a major target for antiviral therapy [7, 8]. In particular, the two classes of antiviral drugs that target RT are nucleoside RT inhibitors (NRIs) and non-nucleoside RT inhibitors (NNRTIs) [4, 9, 10]. Both classes inhibit the RT-associated polymerase activity: the NRTIs compete with the natural dNTP substrate and act as chain terminators; whereas the NNRTIs bind to an allosteric pocket and inhibit polymerization non-competitively [4]. Although, a large number of inhibitors that target RT have been approved for treatment [9, 10], there are many complications that hinder the therapeutic effect of these inhibitors. For instance, NRTIs demonstrated high levels of cellular toxicity [11], whereas the highly potent and less toxic NNRTIs are severely weakened by the rapid emergence of drug-resistant strains [1, 12]. Therefore, tackling the drug-resistant implications associated with NNRTIs could result in less toxic and more potent inhibitors as anti-HIV RT. It is strongly believed that more flexible NNRTIs could be more effective against drug-resistant mutants (4, 12, 13).

The concept of designing flexible RT inhibitors has been proposed before (*4, 12, 13*). Ohtaka and his colleagues made a conclusion that a protease inhibitor required to be flexible in nature in order to bind to mutant enzymes [13]. Soon afterwards, Das *et al* (2005), concluded that conformational flexibility of an NNRTI was crucial in preserving the functionality of an NNRTI being exposed to mutations of resist antiviral strains [14]. Furthermore, it was suggested that the L100 mutant of HIV-1 RT was proficiently inhibited by diarylpyrimidine (DAPY)-based inhibitors [14], as these inhibitors are torsionally flexible and so are capable of binding to any conformation of the hydrophobic pocket of the wild-type RT and other drug-resistant mutants [15].

To this end, in this report, via using integrated *in-silico* approaches, including *de-novo* design, structure-based virtual screening (SBVS) and molecular dynamics simulations, we propose novel flexible hydroquinone-based structures as potential leads to inhibit HIV-1 RT. Specifically, our main purpose is to design and identify flexible ligands with enhanced conformational freedom to allow positional adaptability within RT enzyme active site and as a result could become mutant-resistant inhibitors.

The proposal of hydroquinone-based structures as flexible NNRTIs was first introduced by Bruccoleri [5], however, this proposal was not supported by either experimental or computational evidences. Therefore, the work reported herein is an attempt to verify the activity of hydroquinone-based compounds as NNRTIs, as proposed by Bruccoleri [5]. In order to accomplish this, we have designed and investigated the binding affinity and themes for a novel set of different hydroquinone-based structures against RT enzyme at molecular level. Structure-based virtual screening in order to identify a wider range of novel hydroquinone-based compounds from the commercially available chemical databases then followed this de-novo design. Docking calculations, molecular dynamic stimulations and post-dynamics analysis and per-residue energy contribution were performed in order to accurately estimate the relative binding affinities of the proposed compounds against RT enzyme as well as to ensure the stability of the resulted ligand-enzyme complexes. To validate our calculations, the same docking approach adopted for the proposed compounds was performed for different FDA-approved NNRTIs and results were compared against experimental data.

Information gained from this study could shed light on the activity of a new series of lead compounds as potential anti-HIV RT. Also, the study could be of much interest to medicinal chemistry and biochemistry researchers for further investigation and development.

5.3. Computational Methods

5.3.1. *De-novo* design of the proposed structures and conformational ensembles
Seven proposed structures were modeled, geometry optimized and energy minimized using MMFF94 force field implemented in Avogadros software [16]. Via conformational sampling, low-

energy conformations were produced for each proposed ligand using Omega v1.8b [17]. Conformational isomers developed for each molecule were only examined, if these isomers were within a range of 3.0 kcal/mol from the lowest energy conformer.

5.3.2. Protein systems

The coordinates of the RT enzyme bound to rilpvirine, a new generation NNRTI and commonly abbreviated as TMC 278, were acquired from the protein data bank (http://www.rcsb.org/) (PDB code: 2ZD1). Other NRRTIs used in this study were also obtained from the protein data bank and identified by the following PDB codes: nevirapine (NVP) (PDB code: 4B3Q) and efavirenz (EFZ) (PDB code: 1IKW) (**Figure 5.1.**).

The full computational scenario describing the preparation of enzyme systems for subsequent docking and MD simulations can be found in our earlier reports [18-22].

5.3.3. Docking calculations: Structure-based virtual screening (SBVS).

Docking calculations were conducted using Autodock Vina [23] screening software. The Screening procedure was run using the software default settings; however, the exhaustiveness of the screening was set to the value of 8. The grid box used to define the screening site was elucidated using the Autodock Vina functionality built into Chimera [24]. The box was defined around Asp55, Glu56, Asp79, Arg82, Ile84, Pro85, Val99, Lys108, Val123, Arg141 and Ser169. The X, Y and Z centers were defined as 49.7, -27.3 and 38.11, respectively. The X, Y and Z size dimensions were defined as 42, 38 and 48, respectively. Autodock Vina generated screening results in the pdbqt format. Ten compounds with the best binding energies were selected from each of the two libraries. The docking results of these compounds were visualized the Viewdock feature in Chimera software.

To validate the accuracy of docking method in this work, three FDA-approved drugs with known experimental data (IC_{50}) were seeded in the docking runs and results were compared against experimental data.

5.3.4. 2D shape similarity-based compounds library generation

The best-docked structure resulted from the *de-novo* design (**section 5.3.1.**) was used as template for the generation of a 2D shape similarity-based compounds library retrieved from the Chemicalize database (http://www.chemicalize.org/). The structure of the template compound was drawn using the MarvinSketch Software (Chemaxon, http://wwwchemaxon.com). This was

then used to generate a SMILES code, which was used to query the Chemicalize database. The database was queried for all structures that had greater than 70% shape similarity to the reference template. This results in a total number of 1007 compounds. As explained in our previous report, physicochemical filters were imposed to enhance structures mining process [25]. However, in this study, only compounds with molecular weight between 300 and 550 mg/mol will be selected – this resulted in 303 hits. No other criteria were imposed to ensure the inclusion of the maximum number of compounds. These compounds were downloaded as a single file in the mol2 format. Molegro Molecular Viewer (MMV) software suite [26] was then used to create the individual mol2 files. These files were then converted into the pdbqt format required for Autodock Vina using Raccoon software [27]. Conformational sampling (see **section 5.3.1.**) was also performed for all retrieved compounds before docking into the RT enzyme.

5.3.5. Molecular dynamics (MD) and post-dynamic analysis

The two best docked ligand-enzyme complexes resulted from the SBVS for the *de-novo* designed structures and 2D shape similarity-based compounds library were subjected to MD simulations using Amber software following the procedure explained in earlier report [28]. For analyzing the interactions of post-dynamic protein-ligand complexes, the Ligplot programme *(27)* was used to plot the hydrogen bond and hydrophobic interactions between receptor and ligand atoms within a range of 5 Å. Chimera software [24] was used to visualize and prepare figures for the top ranked molecules. Per-residue energy decomposition was computed using Moldock scoring functions *(25)*.

5.4. Results and Discussion

5.4.1. *De-novo* design plan

First, the enzyme target, RT enzyme, active site was carefully examined in order to explore the nature of the binding site pocket. It was observed that the binding themes differ amongst the crystal structures of various NNRTIs to RT. Specifically, NNRTIs bind to the hydrophobic pocket situated in the p66 subunit of HIV-1 RT [5]. In our study, we determined the active site residues of RT to include: Lys-103, Leu-236, Val-108 and Tyr-190. The NNRTI-binding site also has a high degree of flexibility and as a result can bind NNRTIS of diverse shapes and sizes,

which only opens in the presence of an inhibitor [5, 14]. Some NNRTIs display "butterfly-like" conformations whereas others show "horseshoe" conformations [5, 14].

Based on the study reported by Bruccoleri [5] we proposed seven *de-novo* ligands (**Table 5.1.**) that we anticipate to have potent flexible non-nucleoside inhibitory activity against HIV-1 RT.

Our design initially began with a covalently linked hydroquinone trimer, structure **1** (**Table 5.1.**). Then, in order to provide more flexibility to the structure, we introduced O atom and CH_2 spacers between the hydroquinone rings. After docking, we noticed that these compounds do not optimally fill the enzyme active site pocket as evident from the poor binding affinity (structures **1**, **2** and **3** - **Table 5.1.**). Then we opted to design larger structures by adding phenyl rings on the spacer atoms, CH_2, in order to maximize the hydrophobic interactions between the inhibitors and the active site residues (structure **4**, **Table 5.1.**). As expected, this approach has improved the binding affinity by 1.2 kcal/mol (**Table 5.1.**). We tried a parallel approach to the former one, however, this time by linearly expanding the hydroquinone moieties to create tetramers with and without spacers, compounds **5**, **6**, respectively (**Table 5.1.**). However, these compounds showed modest activity when compared to the prototype, rilpivirine. Adding one more hydroquinone moiety to structure **7** has remarkably improved the binding affinity (**Table 5.1.**).
The reason for this observation is perhaps due to the fact that compound **7** is fully occupying the active site, therefore showed more enhanced binding affinity towards RT.

As evident from **Table 5.1**, all the *de-novo* set 1 designed compounds (**Table 5.1.**), following the design plan proposed by Bruccoleri [5], exhibited lower binding affinity towards RT when compared to rilpivirine (-11.4 kcal/mol) – excluding compound **7** which shows a slightly better binding affinity, however, this is very subtle difference if experimental errors are considered.

From these results, it seems to us that the designed compounds exhibit humble activity as NNRTIs. Many factors might contribute to these findings: (1) the docking method applied might not be accurate enough to provide accurate estimates of the binding energies, (2) our chemical intuition might have drifted away and led to the design in the wrong direction (3) our design have not covered a wider range of structural scaffolds which might have higher binding affinities

towards RT. In order to eliminate the first factor, we opted to validate the docking protocol applied in this work. One of the most commonly used approaches to validate docking results is "cross-validation" where other docking programs are used to validate the data obtained from the original docking software. From our experience, we found that using different software, with different scoring functions, could generate different results and be misleading, therefore, we did not opt to use this validation method. We always believe that the most rational way to validate docking calculations, or even any other computational tool, is to run the calculations on a set of compounds with available experimental data. Results from calculations, are then compared against experimental data for validation. Another reliable approach to validate docking results is to perform docking for the bound inhibitors in the X-ray crystal structure of the inhibitor-enzyme complex, then the RMSD of the docked complex is then compared against the native X-ray structure.

To this end, in this work, we performed docking analyses on three FDA approved NNRTI drugs using our docking methodology (**Figure 5.1.**). The PDB coded structures were obtained from the Protein Data Bank (http://www.rcsb.org/) and experimental IC_{50} values were obtained from the Binding database (http:// www.bindingdb.org/). It is worth mentioning that, we have previously used the same docking approach in several experimental reports and it proved reliable to a great extent [18-22]. As evident from the docking results (**Figure 5.1.**), the docked energies are in great accordance with the experimental IC_{50}. This implies that the docking methods used in this work could be reliable enough to estimate the binding affinities for the proposed compounds.

Figure 5.1. should be here

As mentioned above, and as evident from the humble binding affinities of the de-novo designed compounds, our chemical intuition for the design of potential leads might have failed to build up the right structural scaffolds or at least have not covered a wider range of compounds to probably include those with better binding affinities. To this end, we opted to search for similar compounds in the commercially available chemical database (www.chemicalize.org). Our query of the database was based on 2D shape similarity-based identity to our best-docked structure (compound **7- Table 1**) from the *de-novo* design as a reference template. All compounds in the generated library (see **Methods sections for details**) were then docked into the active site of the RT enzyme. As shown in **Table 5.1**, the top 10 compounds from the SVBS session exhibited

remarkably higher binding energies as compared to Rilpvirine (-11.4 kcal/mol), with binding energies ranging from -12.8 to -11.8 kcal/mol.

Table 5.1 should be here

By looking at the structures retrieved from the database, we found that, however, architecturally similar, none of the compounds retained a hydroquinone fragment. Also, in all structures, a carbon spacer exists between all phenyl rings. Another conclusion can be drawn by looking at **Table 5.1**, the *de-novo* designed library vs. virtually screened library, is that inclusion of methyl groups at different positions on the main structure skeleton has a significant positive impact on the binding affinity. This could be explained based on the fact that the NNRT active site is mostly hydrophobic in nature. Based on the conclusions drawn from **Table 5.1**, we opted to optimize the initially *de-novo* designed leads by inclusion of methyl groups on the skeleton of the structures having a spacer between aromatic rings, structure **2**, **4** and **6**, in order to see whether or not this will enhance the binding affinities. Inclusion of methyl groups on the hydroquinone skeleton, structure **18**, **19** and **20**, has improved the binding affinity by 0.5 - 0.9 kcal/mol (**Table 5.1.**). We assume that inclusion of methyl group has increased the chances of hydrophobic interactions with the hydrophobic residues in the enzyme active site. To verify this assumption, we performed interaction and per-residue binding energy decomposition analysis on the post-dynamic ligand enzyme complexes resulted from the docking calculations (see **section 5.4.1.**).

5.4.2. Molecular Dynamics simulations and post-dynamic analysis

In order to obtain more insight on the stability of the resulted docked complexes, the nature of overall interaction theme between the proposed ligands and the target protein and the specific amino acids involved in the ligand binding, we performed a 2 ns MD simulations followed by extensive post-dynamic analyses on the docked ligand-enzyme complexes from the *de-novo* design and chemical database.

From our previous experience with molecular docking, in many occasions, we experienced that even best docked structures may fly away from the enzyme active site within a few picoseconds of MD simulations. Therefore, we believe that docking calculations that are not validated by relatively a long MD run to ensure the stability of the system might not be reliable. Interestingly, for all the compounds-enzyme complexes, the average RMSD values were below 2.2 Å and the

variation of the potential energies falls within 900 kcal/mol and this is a good indication of the system stability. The PDBs for all docked complexes are provided with the supplementary material.

In order to perform additional lead optimization and expansion phases, it is highly beneficial to have a quantitative estimation of the contribution of individual amino acid residues in the enzyme active site towards ligand binding and we achieved this by the computation of the per-residue interaction MolDock scoring function (*25*). Per-residue interactions for the reference compound (TMC 278) and compounds showed improved affinities with RT, compounds **8**, **7**, and **20** are shown in **Figure 5.2.**

<p align="center">Figure 5.2. should be here</p>

From the per-residue interaction calculations, compounds **7**, **8** and **20** demonstrated remarkable interactions with the active site residues Tyr183, Pro97, Leu102, Leu236, Phe229 and Tyr190 (**Figure 5.2.**). However, to further gain better depiction of these interactions, hydrogen bond and electrostatic interactions maps for the 2n MD averaged complex structures of **8** and **20** with RT (**Figure 5.3.**). These plots were created using Ligplot software (*27*).

<p align="center">Figure 5.3. should be here</p>

As evident from the post-dynamic hydrogen bond and electrostatic interactions (Figure 3A) as well as per-residue analysis (**Figure 5.2.**), the *de-novo* design-based compound 20 showed a strong hydrophobic interaction with Tyr183 and hydrogen bond interactions with His237, Lys103, Gly95 and Gln93 (**Figure 5.4A**). **Table 5.1** and **Figure 5.4.** also showed that compound 8 exhibited its favorable affinity than compound 20 due to the increased hydrophobic interactions with the RT active site residues.

The integrated computational tools adopted in this work has explored the inhibitory landscape of a number of hydroquinone-based polymeric compounds as suggested by Bruccoleri [5] and proved that these compounds, to a reasonable extent, could be potential leads as anti HIV RT. However, the bioavailability profile of these compounds could be problematic since most of them violates Lipinski's rule of five (especially, logP values and the number of hydrogen bond donors and acceptors). Also the toxicity profiles of these compounds need to be experimentally evaluated.

We believe that the computational workflow presented in this study could be a useful tool in the process of design and development of novel leads against biological targets

5.5. Conclusion

In the present study, and guided by the study presented by Bruccoleri [5], we proposed seven *de-novo* hydroquinone -based ligands that we anticipated to have potent flexible non-nucleoside inhibitory activity against HIV-1 RT. However, these de-novo proposed compounds exhibited moderate binding affinities when compared to a rilpivirine. To optimize our de-novo design, we opted to mine for a wider range of similar compounds from the chemical database. However, the chemical database did not retrieve any hydroquinone-based structures, it guided us to optimize our leads. Inclusion of hydrophobic moieties on the hydroquinone skeleton has improved the binding interaction with the RT active site. Despite that fact that proposed compounds exhibited comparable binding affinities as rilpivirine, most of these compounds violate Lipinski's rule of five and expected to have low bioavailability profile. Along with the computational insight provided from this report, further experimental investigations are needed to evaluate the bioavailability as well as the toxicity profiles of these compounds.

5.6. Acknowledgement

5.7. References

1.   Wainberg MA, S. J., Montaner JS, Murphy RL, Kuritzkes DR, and Raffi F. (2005) Challenges for the clinical development of new nucleoside reverse transcriptase inhibitors for HIV infection., *Antiviral Therapy 10*, 13-28.

2.   Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.

3.   Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants., *Retrovirology 9*, 1-11.

4.   Esposito, F., Corona, A., and Tramontano, E. (2012) HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with InnovativeMechanisms of Actions., *Molecular Biology International 2012*, 1-23.

5.   Bruccoleri, A. (2013) Positional Adaptability in the Design of Mutation-Resistant Nonnucleoside HIV-1 Reverse Transcriptase Inhibitors: A Supramolecular Perspective., *AIDS Research and Human Retroviruses 29*, 1-9.

6.   Seckler, J. M., Howard, K.J., Barkley, M.D., and Wintrode, P.L. (2009) Solution Structural Dynamics of HIV-1 Reverse Transcriptase Heterodimer., *Biochemistry 48*, 7646–7655.

7.   Ilina, T., LaBarge, K., Sarafianos, S.G., Ishima, R., and Parniak, M.A. (2012) Inhibitors of HIV-1 Reverse TranscriptaseAssociated Ribonuclease H Activity., *Biology 1*, 521-541.

8.   Harpstead, B. (2005) HIV-1 reverse transcriptase structure, activities, and inhibitors., *Basic Biotechnology MMG 445*, 1-7.

9.   Bauman, J. D., Das, K., Ho1, W.C., Baweja, M., Himmel, D.M., Clark Jr,A.D., Oren, D.A., Boyer, P.L., Hughes, S.H., Shatkin, A,J., and Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design., *Nucleic Acids Research 36*, 5083–5092.

10.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets - Infectious Disorders 2*, 17-32.

11.   Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. . (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses, *Antiviral Research 71*, 256.

12.   Gallant, J. E., Gerondelis, P.Z., Wainberg, M.A., Shulman, N.S., Haubrich, R.H., St Clair, M., Lanier, E.R., Hellmann, N.S., and Richman, D.D. (2003) Nucleoside and

nucleotide analogue reverse transcriptase inhibitors: A clinical review of antiretroviral resistance., *Antiviral Therapy 8*, 489–506.

13. Ohtaka H, V.-C. A., Xie D, and Freire E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease., *Protein Science 11*, 1908–1916.

14. Das K, L. P., Hughes SH, and Arnold E. (2005) Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors., *Progress in Biophys and Molecular Biology 88*, 209–231.

15. Sarafianos S, *e. a.* (2004) Taking aim at a moving target: Designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases., *Current Opinion in Structural Biology 14*, 716–730.

16. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An advanced semantic chemical editor, visualization, and analysis platform", *Journal of Cheminformatics 4*, 1-17.

17. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *ournal of Chemical Information and Modeling 50*, 572-584.

18. Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal of Medicinal Chemistry 46*, 3976-3985.

19. Makatini, M. M., Petzold, K., Sriharsha, S. N., Soliman, M. E., Honarparvar, B., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E., Kruger, H. G., and Govender, T. (2011) Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease, *Bioorganic and Medicinal Chemistry Letters 21*, 2274-2277.

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

nucleotide analogue reverse transcriptase inhibitors: A clinical review of antiretroviral resistance., *Antiviral Therapy 8*, 489–506.

13. Ohtaka H, V.-C. A., Xie D, and Freire E. (2002) Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease., *Protein Science 11*, 1908–1916.

14. Das K, L. P., Hughes SH, and Arnold E. (2005) Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors., *Progress in Biophys and Molecular Biology 88*, 209–231.

15. Sarafianos S, *e. a.* (2004) Taking aim at a moving target: Designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases., *Current Opinion in Structural Biology 14*, 716–730.

16. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An advanced semantic chemical editor, visualization, and analysis platform", *Journal of Cheminformatics 4*, 1-17.

17. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database, *ournal of Chemical Information and Modeling 50*, 572-584.

18. Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal of Medicinal Chemistry 46*, 3976-3985.

19. Makatini, M. M., Petzold, K., Sriharsha, S. N., Soliman, M. E., Honarparvar, B., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E., Kruger, H. G., and Govender, T. (2011) Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease, *Bioorganic and Medicinal Chemistry Letters 21*, 2274-2277.

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, *Bioorganic Chemistry 40*, 19-29.

21. Honarparvar, B., Makatini, M. M., Pawar, S. A., Petzold, K., Soliman, M. E., Arvidsson, P. I., Sayed, Y., Govender, T., Maguire, G. E., and Kruger, H. G. (2012) Pentacycloundecane-diol-Based HIV-1 Protease Inhibitors: Biological Screening, 2D NMR, and Molecular Simulation Studies, *ChemMedChem*.

22. Makatini, M. M., Petzold, K., Alves, C. N., Arvidsson, P. I., Honarparvar, B., Govender, P., Govender, T., Kruger, H. G., Sayed, Y., Jeronimolameira, Maguire, G. E., and Soliman, M. E. (2012) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors, *Journal of Enzyme Inhibition and Medicinal Chemistry*.

23. Trott, O., and Olson, A. J. (2010) Software News and Update 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.

24. 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.

25. Soliman, M. E. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Drug Development Research, in press*.

26. CLCBio.

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

28. 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.
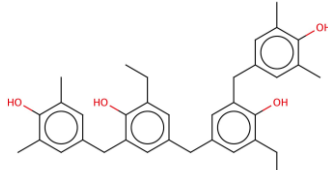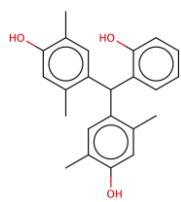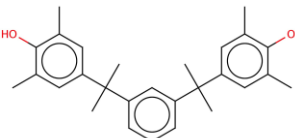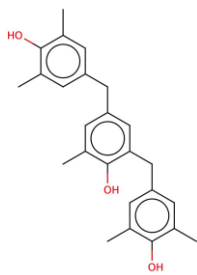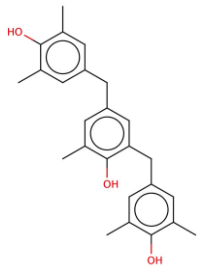
Table 5.1. The 2D structure, calculated binding energy and physicochemical properties of the *de-novo* designed compounds as well as the top-ten best-docked compounds retrieved from chemical database.

Compound	2D structure	Docked binding energy (kcal/mol)	logP	Molecular weight (g/mol)
De-novo set 1				
1		-8.2	2.8	358.3
2		-8.8	3.7	386.1
3		-7.7	2.5	390.3
4		-10.0	3.7	544.6
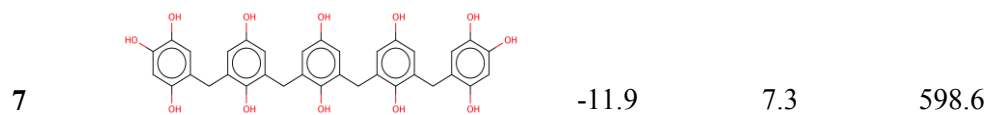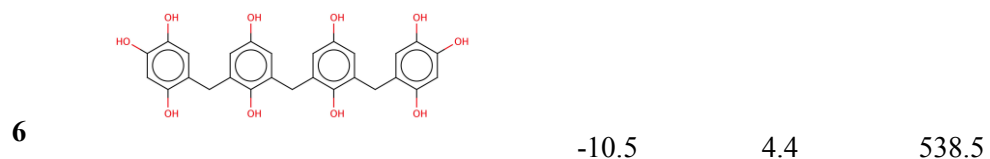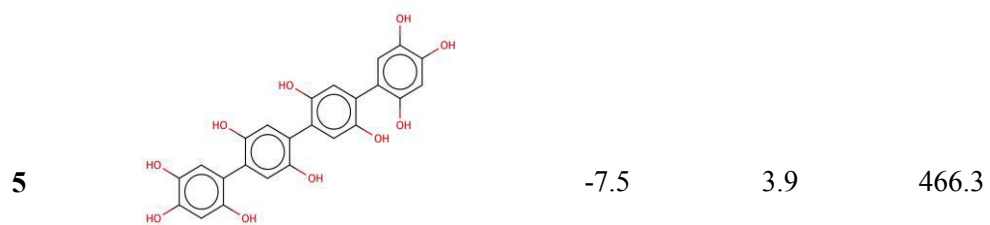
5	-7.5	3.9	466.3
6	-10.5	4.4	538.5
7	-11.9	7.3	598.6

Compounds library from chemical database

8	-12.8	10.11	496.64
9	-12.8	9.09	468.55
10	-12.4	11.00	524.69
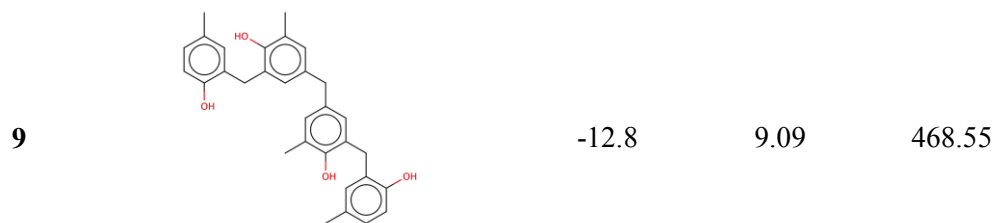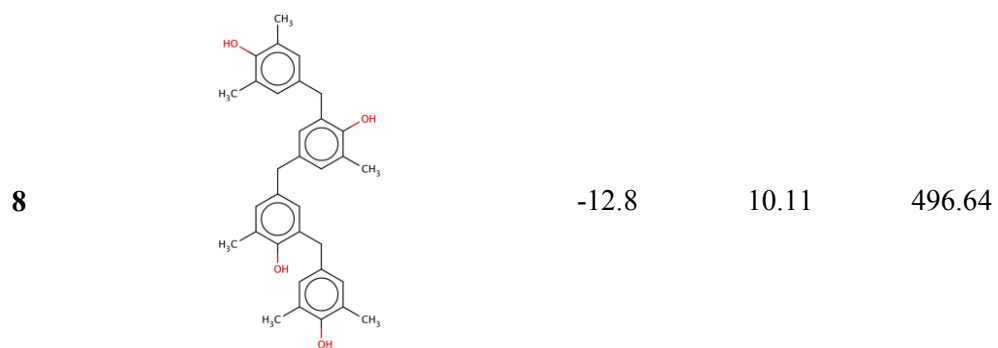
11	-12.3	6.63	348.43
12	-12.2	8.78	402.57
13	-12.0	7.81	376.49
14	-11.9	10.11	496.64
15	-11.8	7.55	424.53

16		-11.8	8.58	452.58
17		-11.8	6.17	340.41

De-novo **set 2**

18		-9.7	5.27	428.43
19		-10.9	8.62	594.23
20		-11.1	7.27	564.58

Figure 5.1. 2D structures of FDA-approved HIV NNRTIs used as reference compounds for validation of docking protocol.

Rilpvirine
(-11.4)*
(421nm)**
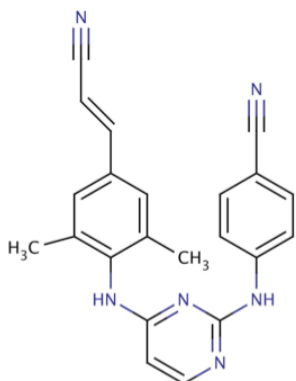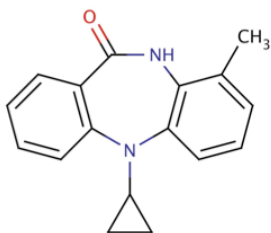
Nevirapine
(-9.1)*
(430nm)**
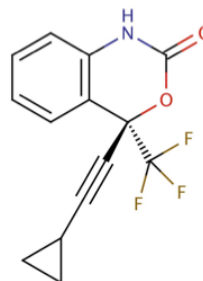
Efavirenz
(-7.4)*
(478nm)**

*Calculated binding energy   **IC$_{50}$

Figure 5.2. Per-residue interactions of the reference compound (TMC 278) and the proposed compounds 7, 8, and 20 with R
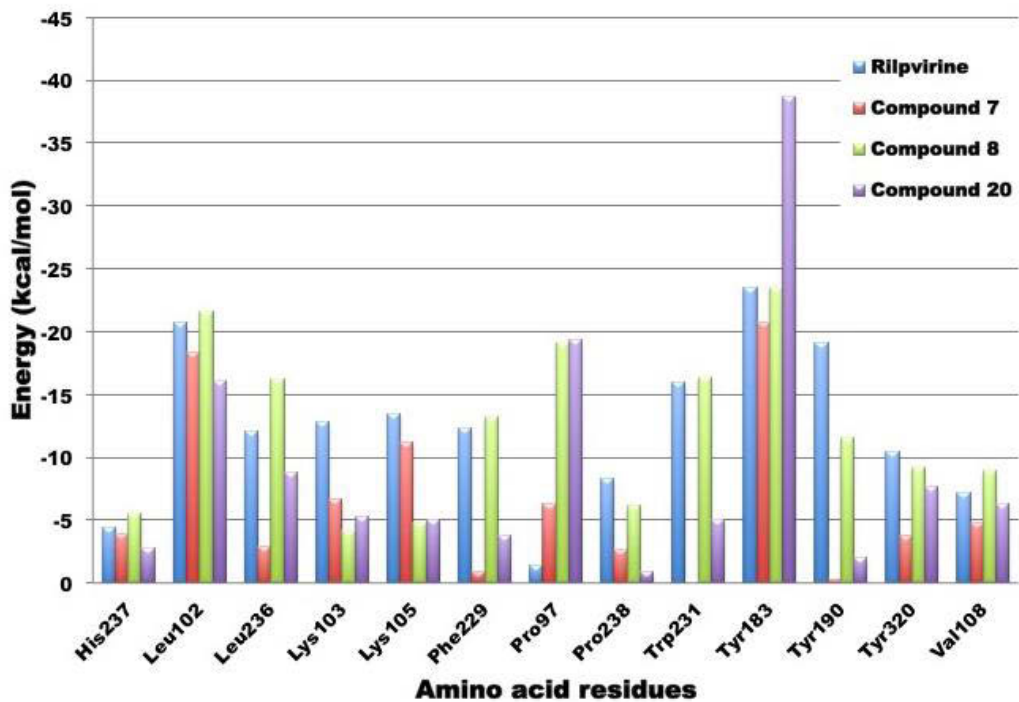
Figure 5.3. Hydrogen bonding and electrostatic interactions for the *de-novo* design-based compound 20 in complex with RT (A) and compound 8 with RT (B), respectively.
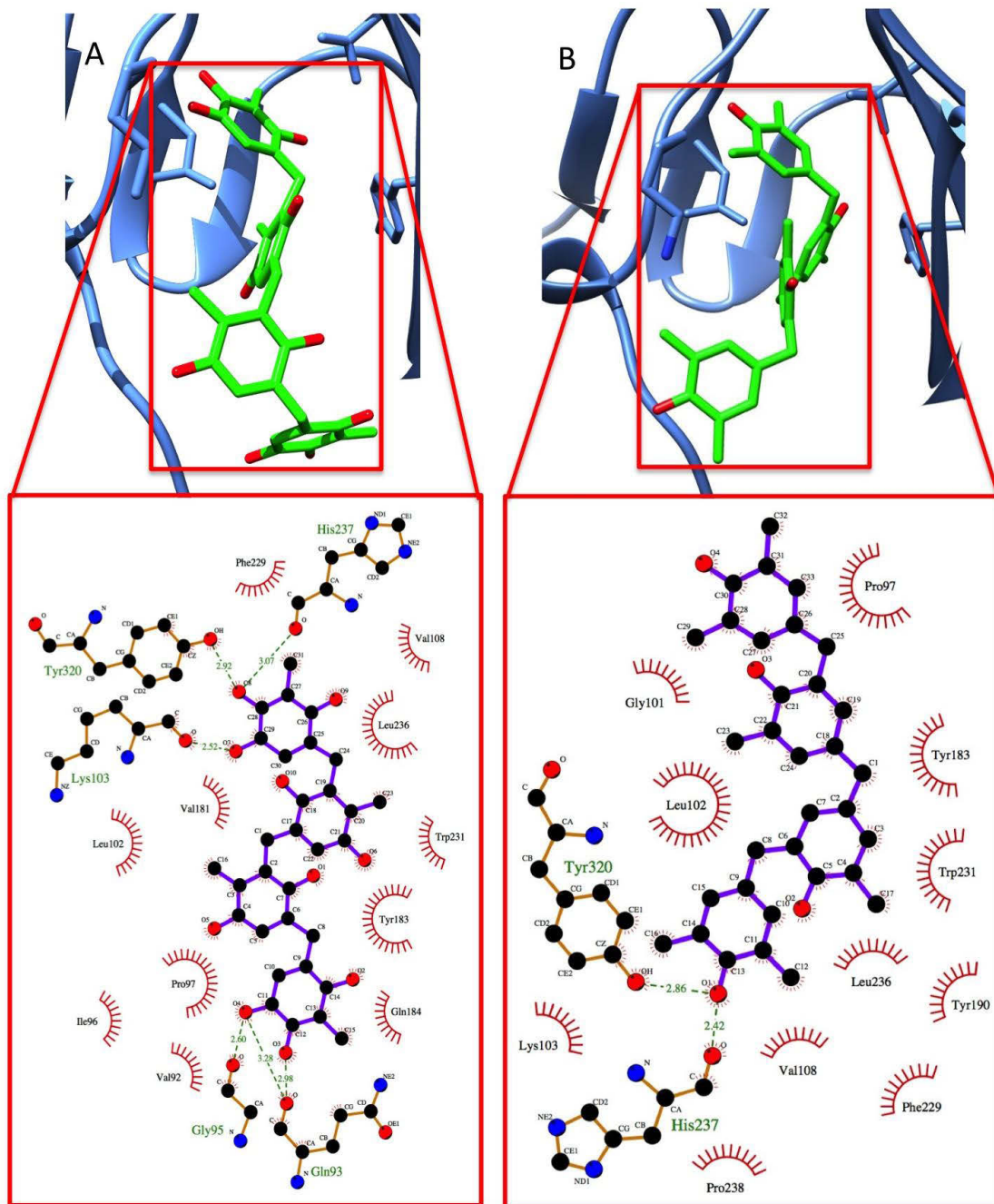
A
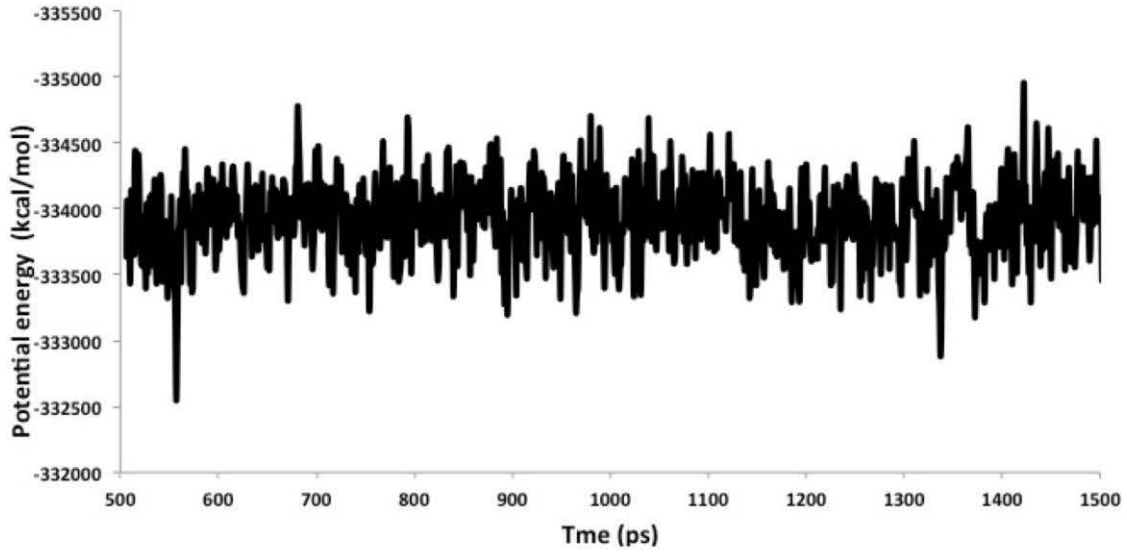
B
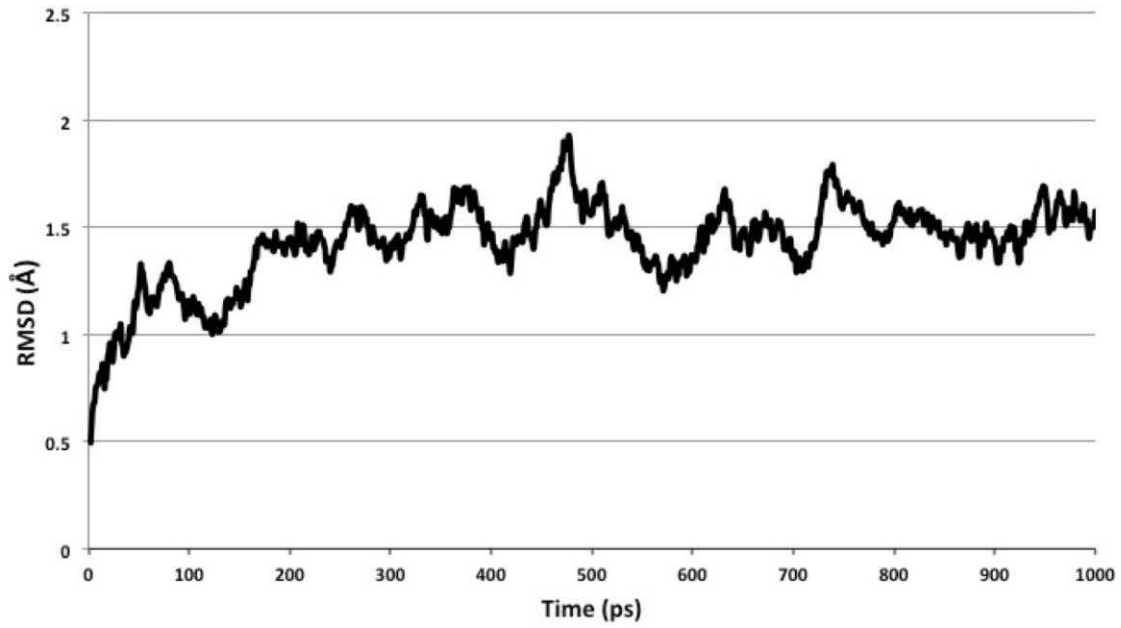
CHAPTER 6

Integrated computational tool for identification of CCR5 antagonists as potential HIV-1 entry inhibitors: homology modeling, virtual screening, molecular dynamics simulations and 3D QSAR analysis

Suri Moonsamy,[a] Radha Charan Dash [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. Soliman, email: soliman@ukzn.ac.za

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

6.1. Abstract

Exploiting an integrated *in-silico* computational technique encompassing homology modeling, structure-based and pharmacophore-based virtual screening, molecular dynamic simulations, per-residue energy decomposition analysis and atom-based 3D-QSAR analysis, we proposed ten compounds as novel CCR5 antagonists as potential HIV-1 entry inhibitors. Via validated docking calculations, binding free energies revealed that the novel leads demonstrated better binding affinities with CCR5 compared to Maraviroc, an FDA-approved HIV-1 entry inhibitor and in clinical use. Per-residue interaction energy decomposition analysis on the averaged MD structure showed that hydrophobic active residues Trp86, Tyr89 and Tyr108 contributed the most to inhibitor binding.

The validated 3D-QSAR model showed a high cross-validated r_{cv}^2 value of 0.84 using three principal components and non-cross-validated r^2 value of 0.941. Also it was found that almost all compounds in the test set and training set yielded a good predicted value.

Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors and serve as a powerful tool in the drug design and development machinery.

Keywords: CCR5 antagonists, HIV-1 entry inhibitors, homology modeling, virtual screening, molecular dynamic simulations, 3D QSAR analysis, computer-aided drug design.

Running title: CCR5 antagonists as potential HIV-1 entry inhibitors.

6.2. Introduction

The human immunodeficiency virus type 1 (HIV-1) infection, which eventually leads to the acquired immunodeficiency syndrome (AIDS) [1], remains to be a fatal disease threatening human health [2]. An estimated 34 million people live with HIV/AIDS worldwide [3, 4]. Approximately 22.9 million individuals in sub-Saharan Africa constitute the overall global estimate [3-5].

In AIDS therapy, the main treatment strategy is to inhibit HIV-1 replication and slow the destruction of the immune system. Currently, a number of viral targets are being used to develop anti-HIV drugs. These targets are mainly viral proteins which are essential for viral replication and survival, for example, protease enzyme (PR) [6], reverse transcriptase (RT) [7] and integrase (IN) [8]. Several drugs which are currently in clinical use have been developed to inhibit these potential viral targets [9].

Numerous concerns regarding the long-term side effects of antiretroviral drugs and the increasing transmission of resistant variants accentuates the requirement to identify new classes of drugs, which are able to suppress HIV-1 replication efficiently [2]. Therefore, there is an on-going need for novel therapeutics, which can prevent the entry of HIV-1 into its target cells [10, 11].

The entry of the HIV virus into its target cell is mediated by the specific interactions of the target cell itself such as the interaction between gp120 viral envelope glycoprotein and the plasmatic membrane receptors [12]. In turn these specific interactions produce conformational alterations in both the glycoprotein and in the membrane receptors that facilitates fusion of the HIV virus and the target cell. Numerous studies have evaluated the role of CD4 and its interaction with gp120 and concluded that the CD4-gp120 interaction is a crucial component, but is not adequate for the disease to become established [13, 14].

Besides CD4, certain chemokine receptors (CCRs) belonging to the G-protein coupled receptor superfamily (GPCRs) have recently been identified as co-targets essential for HIV viral entry into target cells [10, 15]. The different CCRs and their natural chemokine ligands

such as RANTES, MIP-1alpha and MCP2, which are responsible for signaling regulation amidst immune cells and thus signify potential target systems for preventing virus-cell fusion. Several studies have reported on the identification and characterization of diverse CCRs [16]. Besides a CCR that is viral strain dependent, the majority of CCR strains, which are the R5 isolates, transmitted during sexual intercourse and act on CCR5 throughout the disease [16].

The chemokine co-receptor CCR5 has proven to be a critical pharmaceutical target in the contexts of HIV-1 and inflammatory diseases. CCR5 functions as an integral protein in the HIV-1 entry into host cells by acting as a crucial co-receptor for the gp120 viral envelope glycoprotein [1]. Furthermore, the impact of CCR5 in HIV-1 transmission can moreover be demonstrated from experimental data, whereby individuals that are homozygous for the 32-base pair deletion for the CCR5 allele produce a defective CCR5 co-receptor and are resistant to R5-tropic HIV-1 infection, however are otherwise generally healthy [17]. This fact has prompted an enormous research impact in the past decade for identifying anti-HIV agents that specifically target the CCR5-mediated entry mechanism. Furthermore, this implies that functional inhibition of CCR5 may help protect against infection without provoking damage to patients and, therefore, that blocking viral entry using small-molecule antagonists selective for this receptor might provide a new and more effective type of anti-HIV drug.

The concept of designing small molecule CCR5 antagonists has been investigated before [1, 11, 18]. Although no structural information exists about the precise binding site of CCR5 with any FDA approved inhibitor is available. Numerous studies have demonstrated CCR5 modeling of possible inhibitors using computational approaches including virtual screening, molecular docking, molecular dynamic stimulations and pharmacophore modeling. Perez-Nueno *et al* (2008), reported a detailed comparative report of ligand-based and receptor-based virtual screening methods to discover potential HIV entry inhibitors for CXCR4 and CCR5 receptors [19]. In theory, virtual screening of structure-based approaches yields better results than ligand-based methods. Afantitis *et al* (2006) and Aher *et al*. (2007), revealed CCR5 antagonists derived from 1-(3, 3-diphenylpropyl)-piperidinyl amides using virtual screening and quantitative structure-activity relationships (QSAR) studies [20, 21]. In a previous report by

Kellenberger *et al* (2007), structure –based techniques were used to model the physics of protein-ligand interactions in conjugation with combined 2D and 3D structure-based methodologies, which were utilized to find 10 CCR5 binders from a library of 1.6 million compounds [22]. Researchers have developed new approaches of combining computational molecular modeling methodologies, for example molecular docking, 3D-QSAR, comparative receptor modeling and virtual screening to discover potential CCR5 HIV-1 entry inhibitor drugs [1]. Xu *et al* (2004), studied the detailed interactive relationship between CCR5 and its inhibitors using a docking-based 3D-QSAR approach along with protein modeling and MD stimulation [2]. However, in other mechanistic studies of protein-ligand entry inhibitor interactions, investigators have used homology modeling, molecular docking and molecular dynamic stimulation techniques [1].
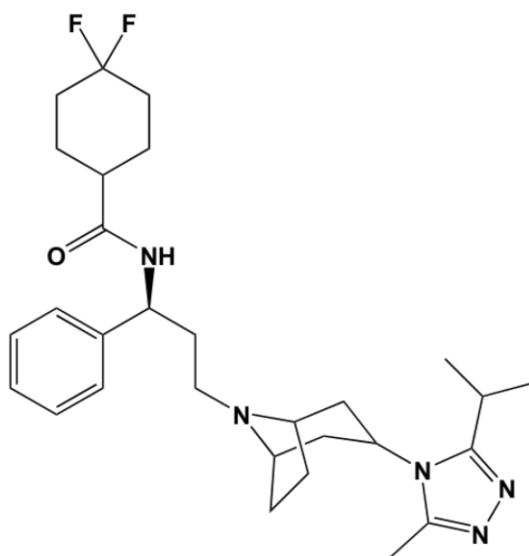
Figure 6.1. 2D Structure of the known FDA-approved CCR5 antagonist, Maraviroc.

To this end, in this report, via a hybrid structure-based and ligand-based virtual screening, we aim to identify novel CCR5 antagonists as potential HIV-1 entry inhibitors based on a human CCR5 homology model template and a known FDA-approved CCR5 antagonist (**Figure 6.1.**), maraviroc, as prototypes. Virtual screening of ligand-based compounds libraries were generated via two distinct yet complimentary approaches: (a) structural similarity-based compound generated library – this library generated compounds that bear a 2D structural similarity to maraviroc, whereas the (b) Pharmacophore-based generated library – this library

generated compounds that contained the pharmacophoric groups of the maraviroc structure. Merging this integrated technique allows us to ensure that our generated compounds library encompasses structural units with diversity, yet with mutual pharmacophoric and structural features. Thereafter, the ligand-based compounds libraries were subjected to virtual screening against the CCR5 enzyme using docking calculations.

To validate our docking calculations, the same docking approach adopted for the ligand-based libraries was performed on a group of compounds with known experimental data attained from inhibition assays against HIV-1 CCR5 and the results were compared against experimental data. Furthermore, from our previous experience with molecular docking, and in numerous insistences, the reliability of a stable protein-ligand complex might not be a true reflection. Therefore, in order to obtain more insight on the stability of the resulted docked complexes, the nature of the overall interaction themes between the generated ligands and the target protein, and the specific amino acids involved in the ligand binding, we performed 1 ns MD simulations followed by extensive post-dynamic analyses on the ligand-enzyme complexes resulted from our docking simulations.

We took our study a step further by obtaining a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC_{50} (mM) data taken from literature for the development of our atom-based 3D-QSAR model [1].

It is worth mentioning that the three dimensional (3D) CCR5 structure is not yet available. However, homology-modeling of CCR5 has been performed before [10]. Therefore, in this study, the actual human CCR5 homology model was developed using the crystal structure of CXCR4 as a structural template. Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors. This study should serve as a powerful tool in the drug design and development machinery.

6.3. Computational methods

6.3.1. Homology modeling of CCR5

In order for our molecular docking study to be executed, the crystal structure of human CCR5 was homology modeled using the human CCR5 protein sequence retrieved from the UniProt database [23] (Uniprot ID: P51681). The actual homology model of CCR5 was created using the CXCR4 crystal structure (PDB ID: 3ODU) as a structural template using the Modeler software [24] add-on in Chimera [25]. Our enzyme model was established with hydrogens, whilst all the important active site amino acid residues numbers were identified using Chimera Multi-align Viewer [25].

6.3.2. Maraviroc structure acquisition and preparation

The structure of maraviroc, an FDA-approved known CCR5 antagonist, in a mol2 file format was acquired from the DrugBank [26-28]. This antagonist had its geometry optimized and energy minimized using MMFF94 force field executed in Avogardros software [29]. Thereafter, for succeeding analyses, the CCR5 antagonist was kept in a mol2 file format.

6.3.3. Ligand library generation

6.3.3.1. Structural similarity-based compound library generation

Maraviroc was used as a template for the generation of a 2D shape similarity-based compounds library generated from the Zinc database. The maraviroc structure was drawn using the MarvinSketch software (Chemaxon, http://www.chemaxon.com). This was then utilized to create a SMILES code, which was employed to query the Zinc Database. This database was queried for all structures that had greater than 60% shape similarity to maraviroc. The query search generated a total of 1002 compound hits. As explained in our earlier report, certain physiochemical filters were implemented to enhance the structure query process [3]. In this study, the pre-defined subset physiochemical filter was set at drug-like qualities. Only compounds with molecular weight between 150 and 500 were selected – this resulted in 480 hits. Other criteria were imposed to ensure the inclusion of the maximum number of compounds, such as the compounds had to have an xlog P between -4 and 5, a net charge between -5 and 5, rotatable bonds between 0 and 8, a polar surface area of between 0 and 150, have hydrogen bond donors/acceptors between 0 and 10, and polar desolvation

between 0 and 1kcal/mol whereas compounds must have an apolar desolvation between -100 and 40kcal/mol. Thereafter, these compounds were downloaded as a single mol2 file format and were individually separated into mol2 files using Molegro Molecular Viewer (MMV) software suite [30]. These files were then converted into pdbqt format needed for Autodock Vina using Raccoon software [31].

6.3.3.2. Pharmacophore-based library generation

The Zinc Pharmer database [32] was used to generate the pharmacophore screening compound libraries. We imported our maraviroc inhibitor structure into Zinc Pharmer with our chosen pharmacophoric regions as illustrated in **Figure 6.2**. These were chosen based on their probability of forming good interactions with the receptor. Furthermore, the Lipinski Rule of Five was imposed as the set criterion for screening compounds. Also, it was imposed for constriction of conformational variations of the same ligand. The purpose of doing this was to reduce any duplication, since the virtual screening software executed diverse conformational variations of the same ligand. The query search generated a total of 602 compounds. All these compounds were downloaded as a single sdf file format and then separated and processed as explained above in **section 6.3.3.1**.
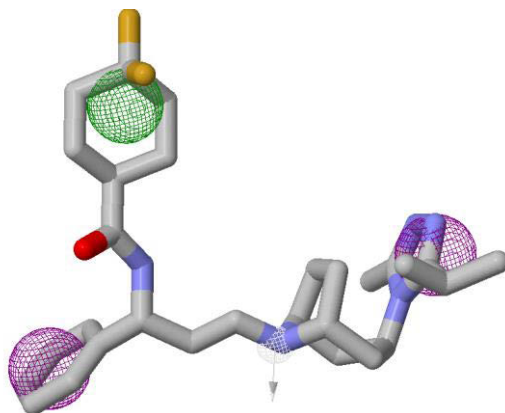
Figure 6.2. (A) Maraviroc structure used as a template for pharmacophore-based and structural similarity-based compound library generation. Pharmacophore selection criteria. Green depicts hydrophobicity, purple depicts aromatic and white depicts hydrogen donor. Arrows indicate that constraints have been imposed.

6.3.4. Virtual screening and validation of docking protocol

In our study, the known CCR5 antagonist (Maraviroc) and the two respective ligand-based compound libraries were subjected to virtual screening against the crystal structure of the human HIV-1 CCR5 enzyme. The Autodock Vina [33] screening software was used to conduct docking calculations. Although the Screening procedure was run using the software default settings; the exhaustiveness of the screening was fixed to the value of 8. The grid box used to define the screening site was clarified by using the functionality property of Autodock Vina, which is constructed in Chimera [25]. The grid box was defined around the following amino acid residues, namely Phe85, Trp86, Trp 89, Leu104, Tyr108, Ile 198, Try251 and Glu283, which resembled the active site residues found in the crystal structure of CXCR4 enzyme following the sequence alignment done in Chimera. The X, Y and Z centres were defined as 11.01, -2.08 and 45, 69, respectively. The X, Y and Z size dimensions were defined as 58, 82 and 74, respectively. Autodock Vinas produced screening results in the pdbqt format. From each of the two compound libraries, we selected ten compounds that displayed the best binding energies. Afterwards, the obtained docking results for each of the ten compounds from each library were visualized using the Viewdock feature in the Chimera software.

6.3.5. Molecular dynamics simulations and post-dynamic analysis

The best-docked ligand-enzyme complexes resulted from the structure-similarity and pharmacophore-based compounds library were exposed to MD stimulations using the Amber software [34] following the procedure explained in our previous report [33]. In order to examine the post-dynamic ligand-protein interactions, in terms of plots representing the hydrogen bond and hydrophobic interactions between the protein and ligand atoms within a range of 5Å were performed using the Molecular Viewing Operator (MOE) program (http://www.chemcomp.com/). The top-ranked were visualized and their figures prepared using the Chimera software [25]. Moldock scoring functions [30] was employed for computing the per-residue energy decomposition calculations.

6.3.6. Three-dimensional (3D) QSAR analysis

A set of 35 novel oxamino-piperidino-piperidine amide analogs (**Figure 6.4.**) with available IC_{50} (mM) data was taken from literature for the development of the atom-based 3D-QSAR model (**Table 6.1.**) [1]. This 3D-QSAR study has been carried out in Discovery studio 3.5

(http://www.accelrys.org). The $1/logIC_{50}$ value of CCR5 was used in this study. Of the 35 compounds reported, 26 compounds were used as a training set and the remaining 9 compounds were used as a test set, based on a random selection. The compounds in the test set have a range of biological activity values similar to that of the training set. The ligands were pre-aligned using a molecular overlay method and placed in a 3D grid space (**Figure 6.3.**). The grid spacing was 1 Å. The energy potentials on every grid point were then calculated using a CHARM force field which used the electrostatic potential and the Van der Waals potential and treated as separate terms. A +1e point charge is used as the electrostatic potential probe and distance-dependent dielectric constant is used to mimic the solvation effect. For the Van der Waals potential a carbon atom with a 1.73 Å radius is used as a probe. The energy grid potentials can be used as independent variables to create partial least-squares (PLS). Further the best 3D-QSAR model was validated by predicting activities of the 9 test set compounds. The 3D-QSAR was evaluated by cross-validated R^2, Q^2. The predicted $1/logIC_{50}$ at 6th PLS factor are tabulated in **Table 6.1**.
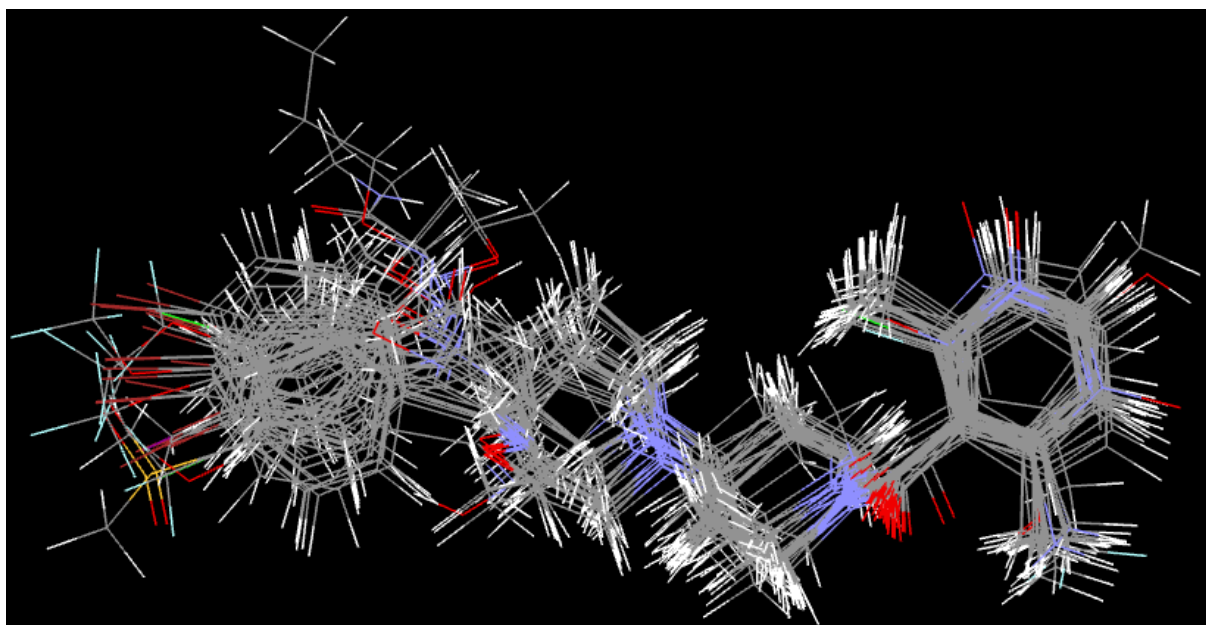
Figure 6.3. Molecular alignments used in the present study.
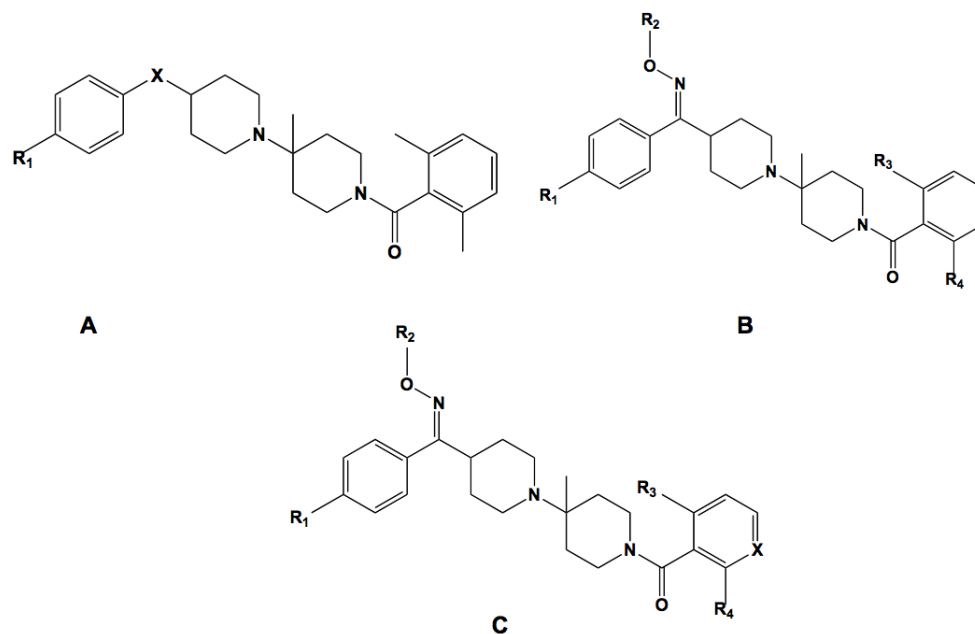
Figure 6.4. The 2D structures for the oxamino-piperidino-piperidine amide analogs used in the 3D QSAR of this work.

Table 6.1. Dataset analyzed for 3D QSAR with experimental $1/\log IC_{50}$, predicted $1/\log IC_{50}$ and residual value.

#	Core	X	R_1	R_2	R_3	R_4	Expt. $1/\log IC_{50}$ (mM)	Prdt. $1/\log IC_{50}$ (mM)	Residual
1	A	CH_2	Br	-	-	-	0.250	0.259	0.009
2	A	NH_2	Br	-	-	-	0.267	0.265	0.002
3	A(E)	=N-	Br	-	-	-	0.290	0.270	0.020
4	A(E)	OCH_3	Br	-	-	-	0.252	0.271	-0.019
5	B	=N-	Br	CH_3	Cl	NH_2	0.360	0.359	0.001
6	B	OCH_3	Br	CH_3	CH_3	OH	0.278	0.276	0.001
7	C	-	CH_3	CH_3	CH_3	CH_3	0.267	0.286	-0.018
8	C	-	Cl	C_2H_5	CH_3	CH_3	0.435	0.392	0.043
9	A	N^+-O	Cl	-	-	-	0.229	-0.248	-0.018
10	A	N	I	-	-	-	0.253	0.230	0.023
11	A	CH_2	CF_3	-	-	-	0.333	0.269	0.064

12	A	CH_2	CH3	-	-	-	0.218	0.233	-0.015
13	A	CH_2	OCH_3	-	-	-	0.227	0.233	-0.006
14	A	CH_2	SO_2CH_3	-	-	-	0.260	0.273	-0.013
15	A	CH_2	Br	-	-	-	0.333	0.313	0.020
16	B(Z)	CH_2	Br	H	CH_3	CH_3	0.250	0.245	0.005
17	B(Z)	$C=CH_2$	Br	C_4H_9	CH_3	CH_3	0.266	0.275	-0.009
18	B	-	Br	CH_2-	CH_3	CH_3	0.274	0.292	-0.016
19	B	-	Br	CO-	F	CF_3	0.301	0.305	-0.004
20	C	N	Br	$NHCH_3$	H	CH_3	0.318	0.289	0.290
21	C	N	Br	C_2H_5	CH_3	CH_3	0.338	0.382	-0.044
22	C	N^+-O	Br	C_2H_5	H	CH_3	0.310	0.307	0.003
23	C	N^+-O	CF_3	C_2H_5	CH_3	CH_3	0.329	0.324	0.005
24	C	N^+-O	OCF_3	C_2H_5	CH_3	CH_3	0.270	0.298	-0.028
25	C	N^+-O	OCF_3	CH_3	CH_3	CH_3	0.307	0.314	-0.007
26	C	N^+-O	Cl	CH_3	CH_3	CH_3	0.324	0.324	-0.006
27	A	C=O	Br	C_2H_5	-	-	0.243	0.261	-0.018
28[t]	B	-	Br	C_2H_5	CH_3	NH_2	0.371	0.385	0.086
29[t]	C	N	Br	-	CH_3	CH_3	0.371	0.329	0.076
30[t]	C	N^+-O	Br	CH_3	CH_3	CH_3	0.301	0.302	0.001
31[t]	A	CH_2	-	CH_3	-	-	0.270	0.277	-0.007
32[t]	A	CH-OH	Br	C_2H_5	-	-	0.205	0.254	-0.54
33[t]	B(Z)	-	Br	-	CH_3	CH_3	0.310	0.271	0.38
34[t]	C	-	CH_3	-	-	-	0.263	0.254	-0.091
35[t]	C	N^+=O	CF_3	C_3H_7	CH_3	CH_3	0.321	0.306	0.015
				CH_3					
				C_2H_5					

t = test set. The conformation of compound denoted in ( ).

6.4. Results and Discussion

6.4.1. Homology modeling of CCR5

The actual homology model of human CCR5 was modeled using the 3ODU crystal structure as a structural modeling template. As outlined in **Figure 6.5**, both these proteins demonstrated good structural similarity in and around the active site residues, with most of the residues having relatively identical locations to each other. The Multi-align Viewer tool

located in Chimera recorded a 42.11% shared similarity between the two proteins' sequences; after modeling, the enzyme model had a zDOPE score of 0.91 with an RMSD of 1.1771 Å. Four differences were observed between the active site sequences of our modeled CCR5 and the 3ODU model template, which included Leu204 (3ODU) corresponding to Leu104 (CCR5), Trp85 corresponding to Trp86, and two residue gaps namely corresponding to Leu213 and Phe113 from 3ODU (**Table 6.2.**). We assumed that these noted differences had very minimal effects in the docking study as result of shared structural similarity between the Leu and Trp residues, respectively. Further investigation is required to verify this assumption. All non-modeled regions were removed from the active site to allow emphasis on all the crucial residues and their importance in the active site of CCR5.

A
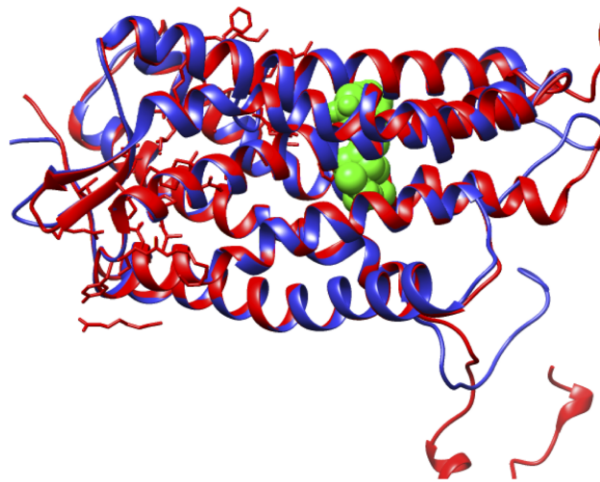
B

```
              1              11             21             31                    41
3ODU (#0) chain A -9 DYKDDDDAGA  PEGISIYTSD  NYTEEMGSGD  YD...SMKEP  CFREENANFN
P51681            1 ..........  ..........  .MDYQVSSPI  YDINYYTSEP  CQKINVKQIA

              51             61             71             81                    91
3ODU (#0) chain A 38 KIFLPTIYSI  IFLTGIVGNG  LVILVMGYQK  KLRSMTDKYR  LHLSVADLLF
P51681           30 ARLLPPLYSL  VFIFGFVGNM  LVILILINCK  RLKSMTDIYL  LNLAISDLFF

             101            111            121            131                   141
3ODU (#0) chain A 88 VITLPFWAVD  AVANWYFGNF  LCKAVHVIYT  VNLYSSVWIL  AFISLDRYLA
P51681           80 LLTVPFWAHY  AAAQWDFGNT  MCQLLTGLYF  IGFFSGIFFI  ILLTIDRYLA

             151            161            171            181                   191
3ODU (#0) chain A138 IVHATNSQRP  RKLLAEKVVY  VGVWIPALLL  TIPDFIFANV  SEADDRYICD
P51681          130 VVHAVFALKA  RTVTFGVVTS  VITWVVAVFA  SLPGIIFTRS  QKEGLHYTCS

             201            211            221            231                   241
3ODU (#0) chain A188 RFYPNDLWVV  VFQFQH...I  MVGLILPGIV  ILSCYCIIIS  KLSHSGSNIF
P51681          180 SHFPYSQYQF  WKNFQTLKIV  ILGLVLPLLV  MVICYSGILK  TL........

             251            261            271            281                   291
3ODU (#0) chain A235 EMLRIDEGLR  LKIYKDTEGY  YTIGIGHLLT  KSPSLNAAKS  ELDKAIGRNT
P51681          222 ..........  ..........  ..........  ..........  ..........

             301            311            321            331                   341
3ODU (#0) chain A285 NGVITKDEAE  KLFNQDVDAA  VRGILRNAKL  KPVYDSLDAV  RRAALINMVF
P51681          222 ..........  ..........  ..........  ..........  ..........

             351            361            371            381                   391
3ODU (#0) chain A335 QMGETGVAGF  TNSLRMLQQK  RWDEAAVNLA  KSRWYNQTPN  RAKRVITTFR
P51681          222 ..........  ...LRCRNEK  K.........  ..........  ..........

             401            411            421            431                   441
3ODU (#0) chain A385 TGTWDAYGSK  GHQKRKALKT  TVILILAFFA  CWLPYYIGIS  IDSFILLEII
P51681          230 ..........  ...RHRAVRL  IFTIMIVYFL  FWAPYNIVLL  LNTFQEFFGL

             451            461            471            481                   491
3ODU (#0) chain A435 KQGCEFENTV  HKWISITEAL  AFFHCCLNPI  LYAFLGAKFK  TSAQHALTSG
P51681          267 .NNCSSSNRL  DQAMQVTETL  GMTHCCINPI  IYAFVGEKFR  NY........

             501            511            521            531                   541
3ODU (#0) chain A485 RPLEVLFQ..  ..........  ..........  ..........  ..........
P51681          308 ..LLVFFQKH  IAKRFCKCCS  IFQQEAPERA  SSVYTRSTGE  QEISVGL
```

Figure 6.5. (A) Superimposed structures of 3ODU [17] and modeled CCR5 enzyme (blue) with CCR5 antagonist, maraviroc [35]. (B) The 2D sequence alignment of 3ODU and the homology model generated for our study. Yellow highlighting represents α-helices and green highlighting represents β-sheets. Sequences outlined in red and black 3D crystal structure.

117
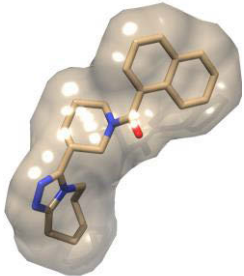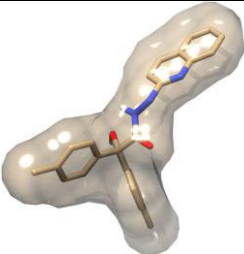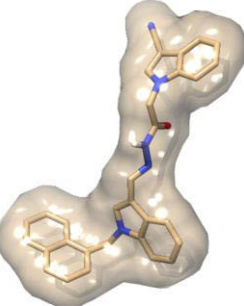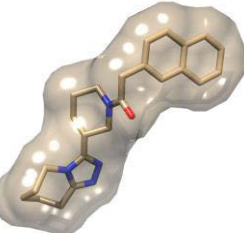
Table 6.2. Comparison of the active site residues between the modeling template (3ODU) and modeled structure.
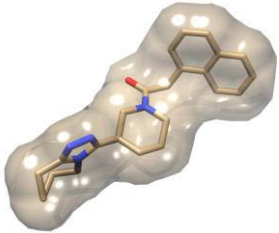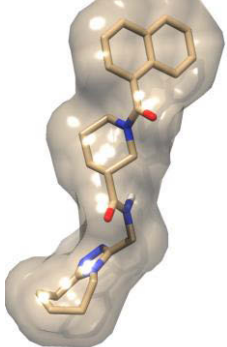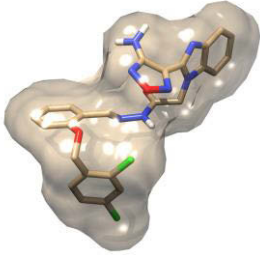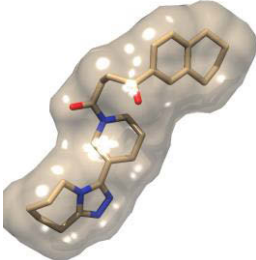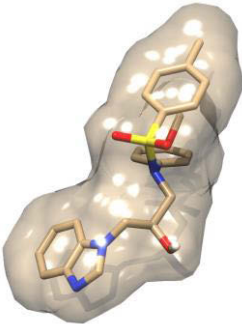
Active site residues (3ODU)	Corresponding modeled active site residues
Glu283	Glu283
Ile198	Ile198
Leu204	Leu104#
Leu213	*
Phe85	Phe85
Phe109	Phe109
Phe113	*
Thr195	Thr195
Trp85	Trp86#
Trp94	Trp94
Trp248	Trp248
Tyr89	Tyr89
Tyr108	Tyr108
Tyr251	Tyr251

#	-	Differing Residue
*	-	Residue Gap

6.4.2. Virtual Screening

Results obtained from virtual screening for the pharmacophore and structure-based libraries are shown in **Table 6.3**.

Table 6.3. List of the top 10 screened compounds based on their docked binding energy. Compounds are ranked in order of highest to lowest binding affinity.

Library	Rank	ZINC ID	Structure	Binding energy	xlogP	H-bond Donors	H-bond Acceptors	Molecular weight (g/mol)
S*	1	ZINC71849549		-12.2	2.27	2	6	318.89
P**	2	ZINC00825224		-12.0	4.11	3	5	397.488
P	3	ZINC00634884		-12.0	5.96	1	6	481.60
S	4	ZINC32760563		-11.9	3.47	0	5	388.51

S	5	ZINC32760533		-11.8	3.44	0	5	388.52
S	6	ZINC25010434		-11.8	2.16	1	7	431.54
P	7	ZINC00851466		-11.8	5.52	3	10	536.38
S	8	ZINC71818945		-11.7	3.35	0	6	434.58
P	9	ZINC00895646		-11.7	3.99	2	7	451.55

P	**10**	ZINC00895634		-11.7	3.54	2	7	438.53
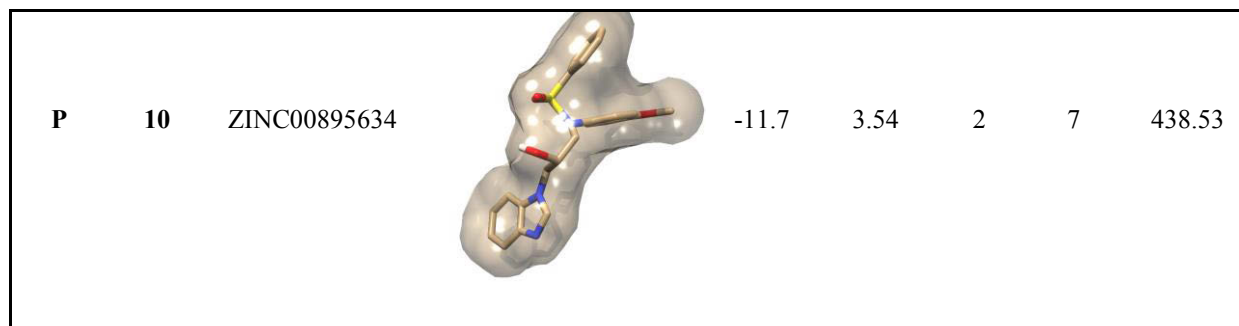
*Similarity-based library, **Pharmacophore-based library

Our query of the ZINC database for compounds sharing 2D shape similarity-based identity to our maraviroc structure used as a reference template generated 220 compound hits. In addition, our query of ZINCPharmer for pharmacologically related compounds generated 120 compound hits. All compounds in the two generated compounds libraries (see Methods sections for details) were then docked into the active site of the CCR5 enzyme using Autodock Vina and thereafter, the top 10 compounds were chosen from each library. As shown in **Table 6.3** and **Figure 6.6**, the entire top 10 ranked compounds from each library exhibited remarkably higher binding energies compared to Maraviroc (-10.2 kcal/mol), with binding energies ranging from -12.2 to -11.6 kcal/mol for the 2D shape similarity-based identity library, whereas the pharmacophore-based library had binding energies ranging from -12.0 to -11.4 kcal/mol.

As outlined in **Figure 6.6**, an unforeseen observation revealed that compounds that were architecturally similar had higher binding energies on average compared to those that were pharmacophorically-based. Although, there was not a huge variance amongst both libraries, however, it raised the inquiry about the particular importance of each pharmacophoric area needed for CCR5 antagonist behavior. Perhaps the reason for this apparent difference is that the chosen pharmacophorically-based vicinities executed significant effects as we originally predicted. A further investigation is required to inquest more intensely into the site specific interactions performed by individual pharmacophoric regions and how does these interactions affect inhibitor binding affinities and functioning. It was also worth noting that compounds 3, 5, 7 and 8 from both the structure-based and pharmacophore-based generated libraries displayed the same binding energies (**Figure 6.6.**). A possible explanation for this observation is that the docking parameters utilized did lead to relatively accurate results. Another interesting observation was that there was not a large difference amongst the top 10 binding energies for

compounds from both the structure-based and pharmacophore-based generated libraries with a difference of -0.60 kcal/mol between the highest and lowest ranked compounds (**Figure 6.6.**). Possibly this observation may be attributed to the conservancy of the crucial pharmacophoric and architecturally shared properties "amidst these compounds allowing for "alleged stability", and hence for well-maintained binding affinities.
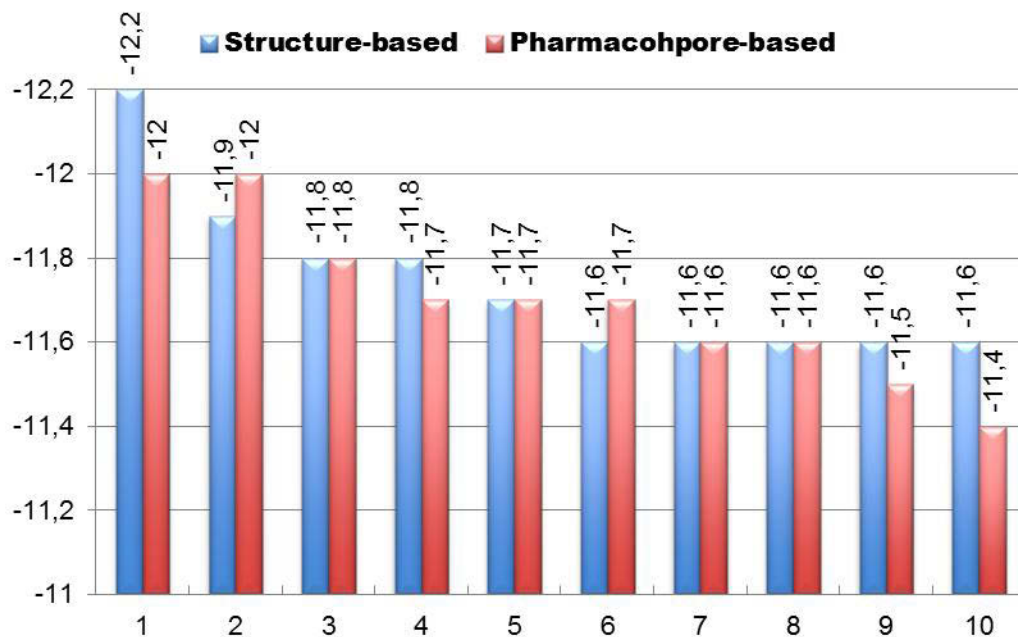
RANK	STRUCTURE-BASED	PHARMACOPHORE-BASED
1	ZINC71849459	ZINC00634884
2	ZINC32760563	ZINC00825224
3	ZINC25010434	ZINC00851466
4	ZINC32760533	ZINC00877991
5	ZINC71818945	ZINC00895634
6	ZINC11546940	ZINC00895646
7	ZINC19698213	ZINC00895774
8	ZINC25010439	ZINC01038388
9	ZINC71849397	ZINC01080881
10	ZINC76956688	ZINC01247854

Figure 6.6. The top 10 ranked ZINC compounds from both the 2D similarity-based and pharmacophore-based libraries.

With the intention of finding the best compounds of this study, the top 10 best-docked compounds from an integration of both compound libraries were revealed (**Figure 6.7.** and **Table 6.3.**). We instantly noticed that these compounds were remarkably larger in size compared to Maraviroc (**Figure 6.7.**). Furthermore, majority of the compounds occupied the spaces between Tyr89, Trp94, Glu283, Leu104, Thr195, Tyr251, Phe109, Ile198 and Trp248, something that was not achieved by Maraviroc as a result of its smaller size (**Figure 6.7.**). It may be deduced that an integration of these factors may be responsible for the higher binding energies demonstrated in this study, as the total number of interactions with the active site seems to be remarkably higher compared to those felt by Maraviroc.
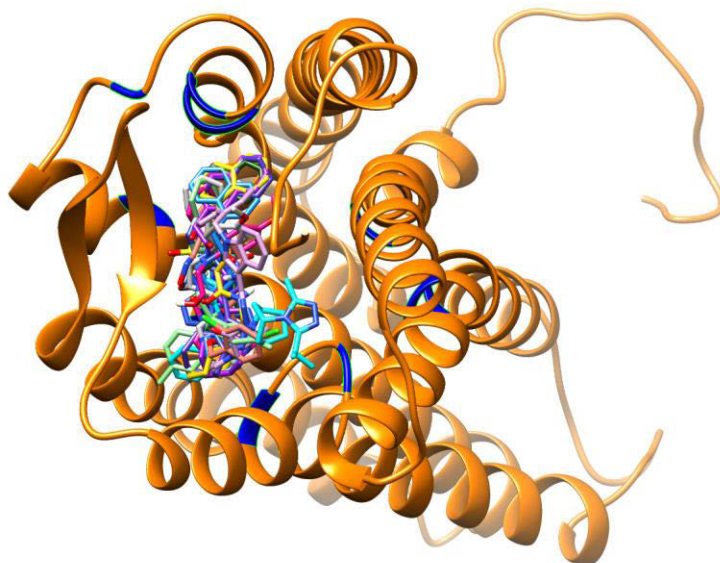
Figure 6.7. Docking conformations of the known CCR5 antagonist (Maraviroc) and the top 10 ranked docked compounds from both the pharmacophore-based and structure-based libraries determined in this study, all complexed with CCR5 enzyme.

From our docking results, it appears that the virtual screening compound hits elucidated exhibit humble activity as CCR5 antagonists. Numerous factors might contribute to these findings, but one of importance to us is that the docking protocol applied might not be accurate enough to provide precise estimates of the different binding energies. In an effort to eradicate this factor, we opted to validate the docking method applied in this study. "Cross validation" was employed to validate our docking results. It is where other docking programs are used to validate the data

attained from the original docking software. Due to previous experience of using different docking software with various scoring functions would generate results that could be different and be misleading, we opted to disregard this docking validation approach. We strongly believe that the utmost rational mannerism for validation of docking calculations, or even any other computational tool, is to perform the calculations on a set of compounds with available experimental data and these results are then compared against known experimental data for validation.

To this end, in this work, we performed docking analyses on a set of compounds assayed using our docking method. The structures and experimental IC_{50} values were obtained from Binding database (http://www.bindingdb.org/). The binding energy of each compound was then plotted against its corresponding experimental IC_{50} value (**Figure 6.8.**). As evident from the docking results (**Figure 6.8.**), the docked energies are in great accordance with the experimental IC_{50}. We observed that as the binding affinity increased (lower binding energy), the IC_{50} increased (**Figure 6.8.**). We postulate that the larger the binding affinity, the more concentration is needed for complete enzyme inhibition. To us, this proves to be an interesting trend, since the top 10 ranked compounds which were obtained from both libraries (**Table 6.3.**) had higher binding energies compared to any of those used in the assay. This implies that the docking approaches used in this work could be reliable enough to estimate the binding affinities for the top 10 ranked compounds (from both libraries – **Table 6.3.**).

	1	2	3	4	5
Binding energy	-11,8	-10,9	-9,1	-8,1	-7,6
IC50	0,2	2,1	10	23	26

Inhibitor
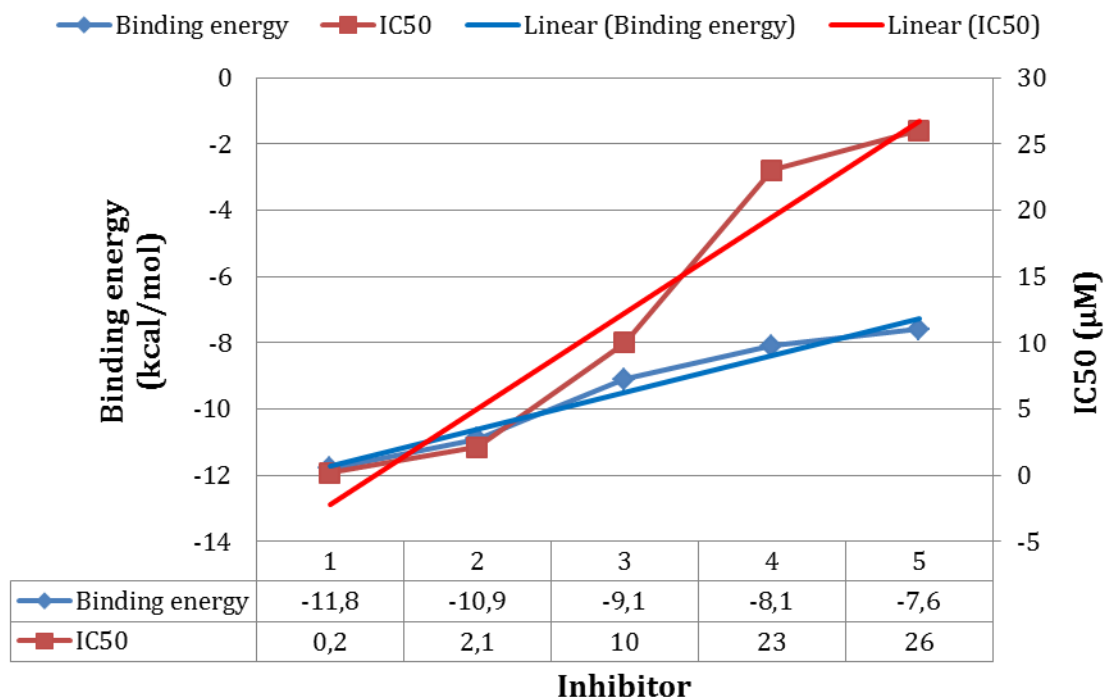
Figure 6.8. The binding energies determined in our study and were compared against IC$_{50}$ values for the compounds assayed. The higher the binding affinity, the higher the IC$_{50}$.

6.4.3. Molecular dynamics simulations and post-dynamic analysis

In order to obtain more insight on the stability of the resulted docked highest ranked virtual screening hit complexes, the nature of overall interaction theme between the proposed ligands and the target protein and the specific amino acids involved in ligand binding, we performed 1 ns MD stimulations followed by extensive post-dynamic analyses on the ligand-enzyme complexes resulted from virtual screening.
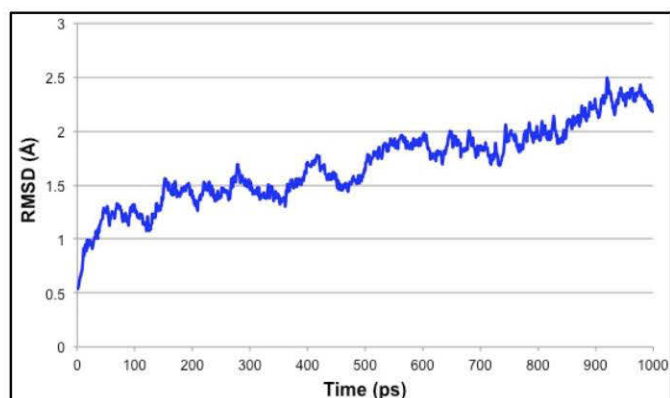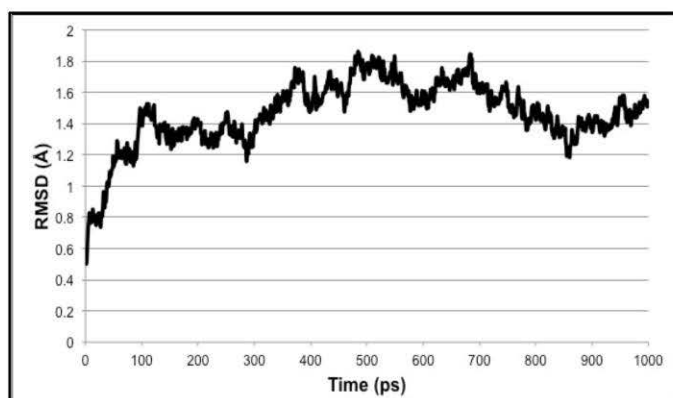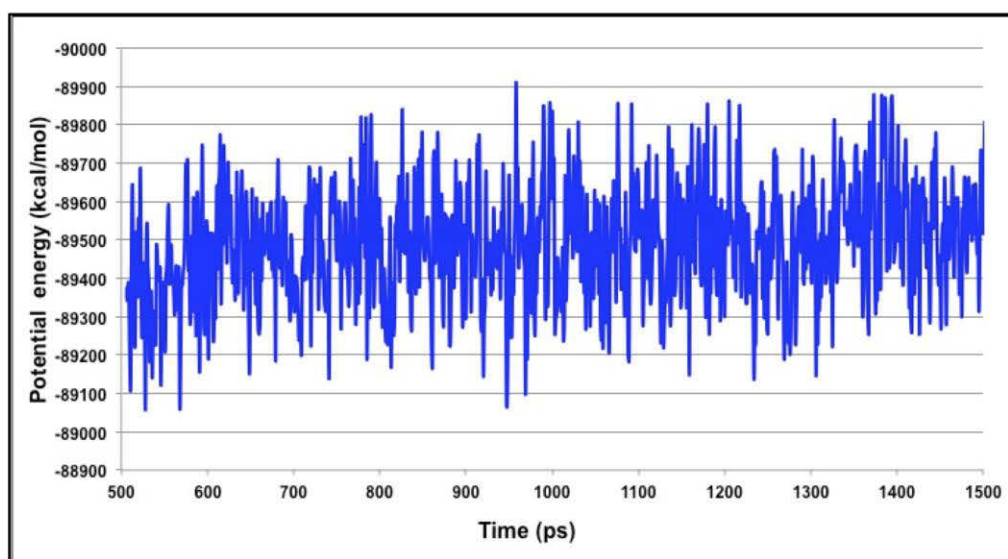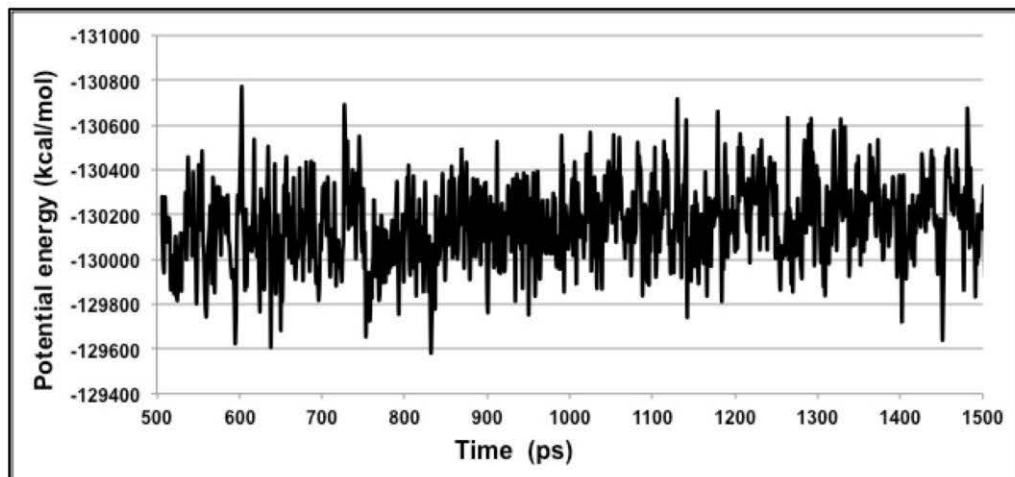
Figure 6.9. The highest ranked virtual screening hit lead complexes with CCR5 subjected to MD simulations. Structure-based compound (ZINC71849459) in complex with CCR5 (Black). Pharmacophore-based compound (ZINC00634884) with CCR5 (Blue).

126

MD stimulations of 1 ns were performed for the highest ranked virtual screening hit complexes (**Figure 6.9.**), to ensure the stability of the ligand within the CCR5 active site. From our previous experience with molecular docking, in many occasions, we experienced that even best docked structures may fly away from the enzyme active site within a few picoseconds of MD stimulations. Therefore, we believe that docking calculations that are not validated by relatively a long MD run to ensure stability of the system might not be reliable. Interestingly, for all the compound-enzyme complexes, the average RMSD values were below 2.5 Å. In addition, the variability of the potential energies fell with 1000 kcal/mol and this suggested being a good indicator of the system stability.

In an effort to investigate the contribution of a single amino acid towards ligand (and/or antagonist) binding, we computed per-residue interactions using Moldock software [30] (**Figure 6.10.**). The top ranked antagonist with the highest binding energy within each compound library was assessed, which included ZINC71849459 from the structure-based identity library and ZINC00634884 from the pharmacophore-based library. We noticed that Lys197, Phe109, Trp86, Tyr89 and Try108 exhibited remarkably interactions with both of those docked ligands. However, Trp86, Tyr89, Thr105, and Tyr108 demonstrated especially better interactions with ZINC00634884 (pharmacophore-based library), whereas Lys197, Leu196 and Glu283 showed good interactions with ZINC71849459 (structure-based identity library). As shown in **Figure 6.11,** in order to gain a better understanding of the ligand-amino acid interactions occurring within CCR5's active site, we generated a plot of the specific amino acid residues-ligand interactions using the MOE software (http://www.chemcop.com/).
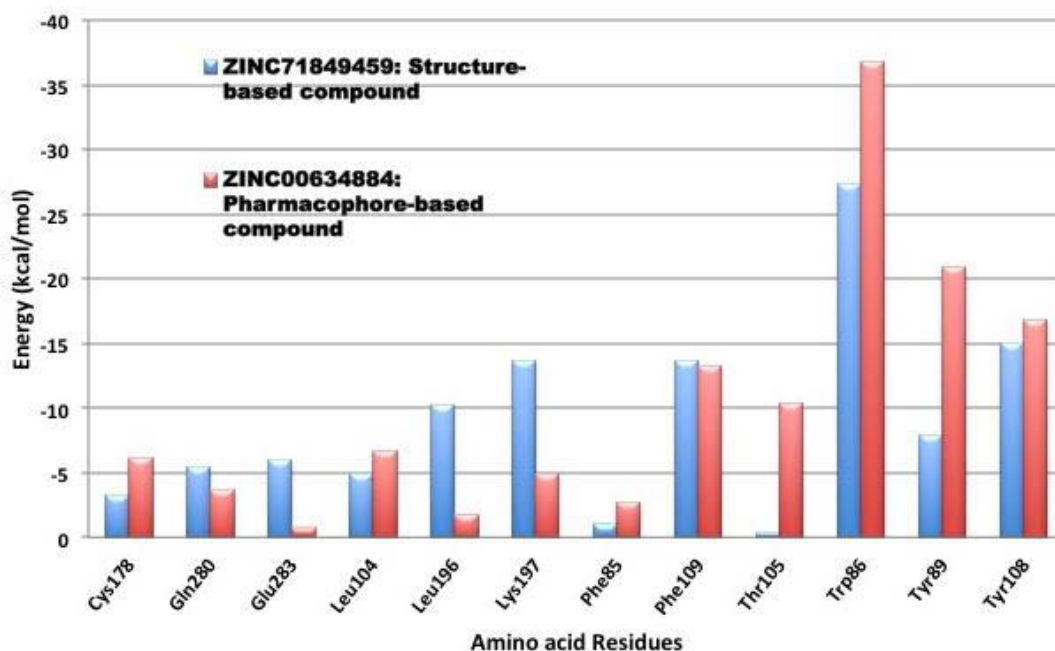
Figure 6.10. Per-residue interactions for the highest ranked compounds with the best binding energy from the structure-based and pharmacophore-based libraries.

The MOE plot analysis of ZINC71849459 bound to CCR5 enzyme's active site revealed that the ligand was especially well surrounded electrostatically by several amino acid residues within the active site (**Figure 6.11a**). However, it could be noted that the ligand ZINC00634884 bound to the active site was not well cradled electrostatically by the amino acid residues compared to the former (**Figure 6.11b**). An interesting observation was that not a single amino acid formed any hydrogen bonds between the protein and ligand for anyone of the two ligands investigated. As shown in **Figure 6.11**, on average the hydrophobic amino acid residues played crucial roles in protein-ligand interactions than the hydrophilic amino acid residues. Another important observation was that the stronger ligand-amino acid residue interactions exhibited here in **Figure 6.11** matched to those outlined in **Figure 6.10**.

A
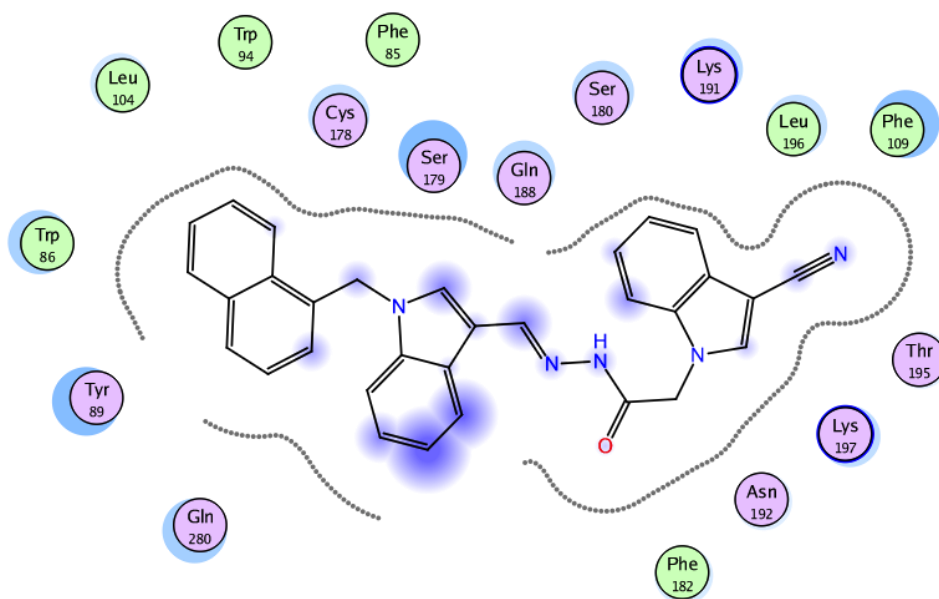
B
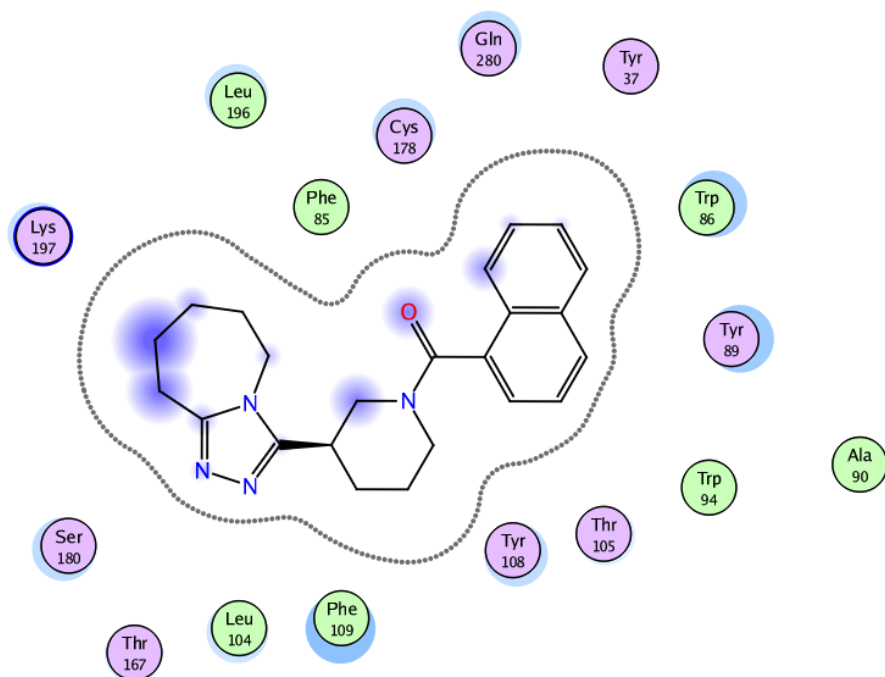
polar
acidic
basic
greasy
proximity contour

sidechain acceptor
sidechain donor
backbone acceptor
backbone donor
ligand exposure

solvent residue
metal complex
solvent contact
metal/ion contact
receptor exposure

arene-arene
arene-H
arene-cation

129

Figure 6.11. Pharmacophore-based compound (ZINC00634884) with CCR5 (A) and Structure-based compound (ZINC71849459) in complex with CCR5 (B), respectively, showing the hydrogen bonding and electrostatic interactions with the enzyme's active site using MOE.

The atom-based 3D-QSAR model developed from the training set of 26 inhibitors (**Table 6.1.**) and the test set of 9 inhibitors using molecular overlay alignments (**Figure 6.4.**). The 3D-QSAR model was built after model development and validation based on the internal predictions of the training set and the external predictions of the test set. PLS analyses of the CCR5 inhibitor training sets showed a high cross-validated r_{cv}^2 value of 0.84 using three principal components and non-cross-validated r^2 value of 0.941. All of the parameters of these QSAR model showed certain reliability and feasible predictability to help us design new and high selectivity CCR5 inhibitors. From **Figure 6.12**, we can see that almost all compounds in the test set and training set yielded a good predicted value. The graphical plot of observed *vs*. calculated TPH1 inhibitory activity for both the training set as well as the test set is shown in **Figure 6.12**.
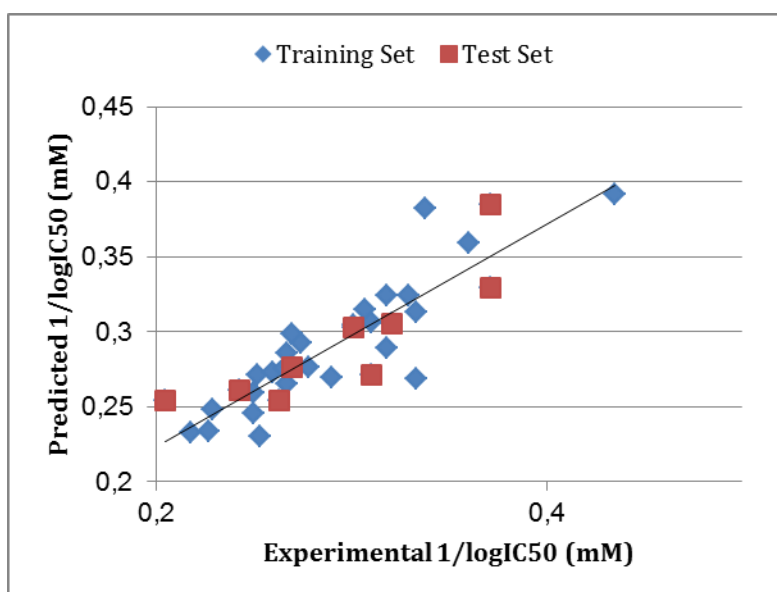
Figure 6.12. Correlation graph between the experimental 1/log IC50 and predicted 1/logIC50.

6.5. Conclusion

In the present work, the structure of human CCR5 was homology modeled using the crystal structure of CXCR4 as a structural modeling template. The protein appeared to be modeled with a remarkably degree of accuracy, specifically at the active site where docking studies were performed. Our query of the ZINC database for drug-like compounds that shared 2D shape similarity-based identity and query of ZINCPharmer for pharmacologically related drug-like compounds to a known FDA-approved CCR5 antagonist called Maraviroc. The entire top 10 ranked compounds from each library exhibited remarkably higher binding energies compared to the best-docked structure of Maraviroc. Moreover, an unforeseen observation revealed that compounds that were architecturally similar had higher binding energies on average compared to those that were pharmacophorically-based. To validate our docking calculations, the same docking approach adopted for the ligand-based libraries was performed on a set of compounds with known experimental data attained from inhibition assays against HIV-1 CCR5 and the results were compared against experimental data. The docked energies are in great accordance with the experimental IC_{50}. We observed that as the binding affinity increased (lower binding energy), the IC_{50} increased. We postulated that the larger the binding affinity, the more concentration is needed for complete enzyme inhibition. To us, this proves to be an interesting trend, since the top 10 ranked compounds which were obtained from both libraries had higher binding energies compared to any of those used in the biological assay, therefore it stands to reason compound hits elucidated in this study exhibit humble activity as CCR5 antagonists. Also, this implies that the docking approaches used in this work could be reliable enough to estimate the binding affinities for other compounds to be studied. Furthermore, from our previous experience with molecular docking, and in numerous insistences, the reliability of a stable protein-ligand complex might not be a true reflection. Therefore, in order to obtain more insight on the stability of the resulted docked complexes, the nature of the overall interaction themes between the generated ligands and the target protein, and the specific amino acids involved in the ligand binding, we performed 5 ns MD simulations followed by extensive post-dynamic analyses on the ligand-enzyme complexes resulted from our docking simulations. We took our study a step further by obtaining a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC_{50} (mM) data taken from literature for the development of our atom-based 3D-QSAR model. All of the parameters of the QSAR model showed certain reliability and

feasible predictability to help us design new and high selectivity CCR5 inhibitors. Our novel identified leads have the propensity to be considered as potential CCR5 antagonists and moreover as potential HIV-1 entry inhibitors. Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors and should serve as a powerful tool in the drug design and development machinery.

6.6. Acknowledgments

6.7. Author Contributions

SM: virtual screening, molecular dynamics of all compounds as well as write up; RCD: 3D-QSAR & write up; MES: principal investigator and corresponding author.

6.8. Conflicts of Interest

The authors declare no conflict of interest.

6.9. References

1.   Gadhe, C. G., Kothandan, G and Cho, S.J. (2012) Computational modeling of human coreceptor CCR5 antagonist as a HIV-1 entry inhibitor: using an integrated homology modeling, docking, and membrane molecular dynamics simulation analysis approach., *Journal of Biomolecular Structure and Dynamics*, 1-26.

2.   Xu, Y., Liu, H., Niu, C. Y., Luo, C., Luo, X. M., Shen, J. H., Chen, K. X., and Jiang, H. L. (2004) Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-yl)-butanes based on the structural modeling of human CCR5 receptor, *Bioorganic & Medicinal Chemistry 12*, 6193-6208.

3.   Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Drug Development Research 74*, 283-295.

4.   Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. . (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants, *Retrovirology* 1-23.

5.   UNAIDS. (2011) World AIDS day report 2011, *http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/jc2216_worldaidsday_report_2011_en.pdf*, 11 Nov. 2013.

6.   Patel, J. R., and Prajapati, L. M. (2013) Predictive QSAR modeling on tetrahydropyrimidine-2-one derivatives as HIV-1 protease enzyme inhibitors, *Medicinal Chemistry Research 22*, 2795-2801.

7.   Zhan, P., Chen, X., Li, D., Fang, Z., De Clercq, E., and Liu, X. (2013) HIV-1 NNRTIs: structural diversity, pharmacophore similarity, and impliations for drug design, *Medicinal Research Reviews 33*, E1-E72.

8.   Johnson, B. C., Metifiot, M., Ferris, A., Pommier, Y., and Hughes, S. H. (2013) A Homology Model of HIV-1 Integrase and Analysis of Mutations Designed to Test the Model, *Journal of Molecular Biology 425*, 2133-2146.

9.   Pani, A., Loi, A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. (2002) Targeting HIV: Old and new players, *Current Drug Targets - Infectious Disorders 2*, 17-32.

10.   Fano, A., Ritchie, D. W., and Carrieri, A. (2006) Modeling the structural basis of human CCR5 chemokine receptor function: From homology model building and molecular dynamics validation to agonist and antagonist docking, *Journal of Chemical Information and Modeling 46*, 1223-1235.

11.   Manikandan, S., and Malik, B. K. (2008) Modeling of human CCR5 as target for HIV-I and virtual screening with marine therapeutic compounds, *Bioinformation 3*, 89-94.

12. Cormier, E. G., Persuh, M., Thompson, D. A. D., Lin, S. W., Sakmar, T. P., Olson, W. C., and Dragic, T. (2000) Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120, *Proceedings of the National Academy of Sciences 97*, 5762-5767.

13. Farzan, M., Vasilieva, N., Schnitzler, C. E., Chung, S., Robinson, J., Gerard, N. P., Gerard, C., Choe, H., and Sodroski, J. (2000) A tyrosine-sulfated peptide based on the N terminus of CCR5 interacts with a CD4-enhanced epitope of the HIV-1 gp120 envelope glycoprotein and inhibits HIV-1 entry, *Journal of Biological Chemistry 275*, 33516-33521.

14. Cormier, E. G., Tran, D. N., Yukhayeva, L., Olson, W. C., and Dragic, T. (2001) Mapping the determinants of the CCR5 amino-terminal sulfopeptide interaction with, *Journal of Virology 75*, 5541-5549.

15. Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major, *Science 270*, 1811-1815.

16. Pease, J., and Horuk, R. (2012) Chemokine Receptor Antagonists, *Journal of Medicinal Chemistry 55*, 9363-9392.

17. Ditzel, H. J., Rosenkilde, M. M., Garred, P., Wang, M., Koefoed, K., Pedersen, C., Burton, D. R., and Schwartz, T. W. (1998) The CCR5 receptor acts as an alloantigen in CCR5Delta32 homozygous individuals, *Proceedings in the National Academy of Science U S A. 95*, 5241-5245.

18. Kothandan, G., Gadhe, C. G., and Cho, S. J. (2012) Structural Insights from Binding Poses of CCR2 and CCR5 with Clinically Important Antagonists: A Combined In Silico Study, *Plos One 7*.

19. Perez-Nueno, V. I., Ritchie, D. W., Rabal, O., Pascual, R., Borrell, J. I., and Teixido, J. (2008) Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking, *Journal of Chemical Information and Modeling 48*, 509-533.

20. Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., and Igglessi-Markopoulou, O. (2006) Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques, *Journal of Computer-Aided Molecular Design 20*, 83-95.

21. Aher, Y. D., Agrawal, A., Bharatam, P. V., and Garg, P. (2007) 3D-QSAR studies of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas as CCR5 receptor antagonists, *Journal of Molecular Modeling 13*, 519-529.

22. Kellenberger, E., Springael, J.-Y., Parmentier, M., Hachet-Haas, M., Galzi, J.-L., and Rognan, D. (2007) Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening, *Journal of Medicinal Chemistry 50*, 1294-1303.

23. Consortium, T. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt), *Nucleic Acids Research 40*, 71-75.

24. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007) Comparative protein structure modeling using MODELLER, *D - 101287868 John Wiley & Sons*.

25. 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.

26. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Research 39*, D1035-D1041.

27. Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Research 36*, D901-D906.

28. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Research 34*, D668-D672.

29. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *Journal of Cheminformatics 4*.

30. Thomsen, R., and Christensen, M. H. (2006) MolDock: a new technique for high-accuracy molecular docking, *Journal of Medicinal Chemistry 49*, 3315-3321.

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

32. Koes, D. R., and Camacho, C. J. (2012) ZINCPharmer: pharmacophore search of the ZINC database, *Nucleic Acids Research 40*, W409-W414.

33. 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 & Drug Design 81*, 208-218.

34. 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.

35. 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.

CHAPTER 7

Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine: Insight from Multiple Molecular Dynamics Simulations

Suri Moonsamy[a,1], Soumendranath Bhakat[a,1], Ross C. Walker[b] and Mahmoud E. S. Soliman[a*]

[a]School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

[b] San Diego Supercomputer Center & Department of Chemistry and Biochemistry, University of California, 9500 Gilman Drive, La Jolla, CA, 92093-0505, USA.

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

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

[1] Authors contributed equally

7.1. Abstract

Molecular dynamics simulations, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analysis were employed in order to investigate the atomistic basis of the mystery of why the M184I single mutation leaves the HIV-1 reverse transcriptase (RT) totally resistant to lamivudine. Results showed that single mutations at residue 184 of RT caused; (1) distortion of the orientation of lamivudine in the active site due to the steric conflict between the oxathiolane ring of lamivudine and the side chain of beta-branched amino acids Ile at position 184 which, in turn, perturbs inhibitor binding, (2) decrease in the binding affinity by (~8 kcal/mol) when compared to the wild type, (3) variation in the overall enzyme motion as evident from the PCA for both systems and (4) distortion of the hydrogen bonding network and atomic interactions with the inhibitor.

The comprehensive analysis presented in this report can provide useful information for understanding the drug resistance mechanism against lamivudine. The results can also provide some potential clues for further design of novel inhibitors that are less susceptible to drug resistance.

Running title:

HIV Reverse transcriptase resistance

7.2. Introduction

Since Acquired Immunodeficiency Syndrome (AIDS) was first reported in the early 1980s, this global killer has claimed the lives of an estimated 22 million individuals [1]. Amongst the classified infectious diseases, the Human Immunodeficiency Virus type 1 (HIV-1) infection, the causative agent of AIDS, remains a challenging epidemic [2]. It has been reported that 34 million people live with HIV/AIDS globally and in sub-Saharan Africa, approximately 22.9 million individuals contribute to the overall global estimate [3].

Despite extensive ongoing HIV/AIDS research, to date, no cure has been found for this disease. Currently, the most effective treatment comprises a complex "cocktail" of Food and Drug Administration (FDA)-approved and clinical trial drugs that involve multiple drug targets [4]. These drugs target and attack the virus at different stages of its lifecycle, thereby halting virus replication and reducing destruction of the immune system. They include: protease inhibitors (PIs), reverse transcriptase (RT) inhibitors, integrase (IN) inhibitors and entry inhibitors [5].

HIV-1 is a retrovirus that replicates within a host cell. The RT enzyme is an essential component for HIV-1 replication [6]. RT is responsible for converting single-stranded RNA viral genome into double-stranded DNA [6, 7]. Because of its integral role in the virus lifecycle, HIV-1 RT is a major target for drug therapy [8].

Currently, approved HIV-1 RT inhibitors used in antiretroviral therapy can be divided into two classes: (a) nucleoside analog RT inhibitors (NRTIs) [9, 10], which compete with natural nucleoside substrates for binding to the RT polymerase active site and after their incorporation into the primer site, they act as terminators of proviral DNA synthesis; (b) non-nucleoside RT inhibitors (NNRTIs), which bind to a hydrophobic pocket close to the RT active site [7, 9]. More specifically when NRTIs are phosphorylated intracellularly, these drugs become activated and hence inhibit reverse transcription by triggering chain termination once incorporated into viral DNA [7, 10].

The initial introduction of NRTIs for treatment of HIV and later presented as the favored central core drugs of highly active antiretroviral therapy (HAART), has substantially reduced HIV-related morbidity and mortality rates [11]. Following zidovudine (AZT) as the first NRTI drug approved by the FDA for HIV therapy [11], numerous other FDA-approved NRTIs have been

discovered including: lamivudine (3TC), stavudine (D4T), tenofovir (TFV) and emtricitabine (FTC) amongst others [11, 12]. However, even though NRTIs have served as the cornerstones of successful HIV therapy, the largest problem in HIV drug therapy is that the virus mutates very quickly, leading not only to drug resistance [13], but also drug-resistant variants of the virus that have mutations in the RT target protein [14, 15]. For this reason a molecular understanding of the impact of mutations on drug resistance will assist in the design of more potent drugs that are active against resistant strains.

An important constituent of triple-drug anti-AIDS therapy is the NRTI 2', 3'-dideoxy-3'-thiacytidine (3TC, lamivudine). It has been proven that single mutations at residue 184 of RT in HIV cause high-level resistance to 3TC and contribute to the failure of anti-AIDS combination therapy [16].

Partially due to the lack of an X-ray crystal structure of the lamivudine-RT complex, the precise mechanism by which the M184I mutant develops resistance to lamivudine is not fully understood. Based on a previously built model from "unbound" X-ray crystal structures of the wild type and M184I mutant, it is thought that steric conflict between the oxathiolane ring of lamivudine and the side chain of beta-branched amino acids Ile at position 184 perturbs inhibitor binding, leading to a reduction in incorporation of the analog [16]. However, this assumption was based on approximated "fitted" and "static" models.

In this work we aim to provide more comprehensive insight into the precise impact of the M184I mutation on resistance to lamivudine. We believe that, to develop potent and effective anti-HIV NRTIs against the viral variants, understanding of molecular basis of the M184I RT mutation at atomic level will be very critical.

In recent years molecular modeling tools have become close counterparts of experiments. In one of our recent papers, the interface between computation and experiment within the biological context was comprehensively highlighted [17]. Various molecular modeling approaches such as molecular docking, molecular dynamics (MD) simulation and binding free energy calculations have proven useful in understanding the molecular basis of drug resistance to different biological

systems. This includes, but is not limited to, the drug resistance mechanisms against HIV protease [18-20], HIV reverse transcriptase (RT) [21], HIV integrase (IN) [22], influenza neuraminidase (NA) [23, 24], and HCV protease [22, 25, 26].

Standard MD analysis for large biomolecular systems can only provide limited information about the dynamic landscape of these systems. During the last few years, multiple approaches have been devised to deal with the large number of explicit degrees of freedom [27-30]. Principal Component Analysis (PCA) or essential dynamics analysis is one of widely used approaches to explore the structural fluctuations of biological systems. PCA is a statistical procedure based on covariance analysis that can transform the original space of correlated variables into a reduced space of independent variables and is often used to highlight large correlated motions in macromolecules [31, 32]. It takes the trajectory of an MD simulation into account and reduces the dimensionality of the data to extract dominant modes in the motion of a molecule [33-36]. In normal mode analysis these pronounced motions resulting from the protein movement correspond to correlated vibrational nodes or collective group motion of atoms [37, 38]. The eigenvector for a given motion has a corresponding eigenvalue, which signifies the energetic contribution of a particular component to the motion. Projection of a trajectory onto a specific eigenvector highlights the time dependent motions that the components perform during a particular vibrational mode [34].

PCA has been used to identify the overall motion of the enzyme backbone to check the difference from one state to another with high utility in the comparison of wild type proteins with those of mutant/resistant strains [39]. Via PCA, it is possible to identify the dominant motions observed during a simulation through visual inspection. A large portion of overall fluctuation of macromolecules can often be accounted for by a few low frequency eigenvectors with high eigenvalues [40]. If motion between two different macromolecules is similar then the eigenvectors coming from individual trajectories should be similar to each other. For this reason, PCA has proved an efficient tool that can be used to compare motions of different macromolecules [41].

It has also been observed that the resistant mutation or substitution in the target protein often impacts the binding ability of drugs through modification of the residue interaction network

(RIN) communication of the targets [42-45]. Recently, some reports have shown that the analysis and interactive visualization of RIN of proteins can reveal important information on biological interactions in complex systems [46, 47].

In this work, using validated models of lamivudine-RT complexes, wild type and the M184I mutant, MD simulations, post-dynamics analysis, binding free energy calculations, PCA and RIN analysis were employed to provide an insight into the drug resistance mechanism of the RT-M184I mutant to lamivudine. To our knowledge, this report is the first account of such comprehensive computational analysis being used to reveal the precise impact of the M184I mutation on RT resistance to lamivudine.

The compilation of the computational and molecular informatics tools presented in this study could serve as powerful tools to understand protein structures and dynamics, drug resistance and could also be incorporated in the drug discovery and development workflows.

7.3. Computational Methods

7.3.1. System Preparation
With the lack of a Lamivudine-RT complex crystal structure, the construction of a reasonable initial set of 3D coordinates for the complex presented a challenge. Since molecular docking can, in many instances, produce unrealistic ligand orientations in the active site, we opted not to perform docking calculations but instead to use a "ligand fitting" strategy based on known X-ray crystal structures of similar inhibitors in complex with RT. This approach has proven successful in the past, as reported previously [48, 49]. The X-ray crystal structure of RT in complex with zidovudine, AZT, (PDB code: **1RTD**) was used as the starting coordinates (**Figure 7.1.**). This choice was based on the fact that since AZT and 3TC are structurally analogous, similar binding modes would be anticipated (**Figure 7.1.**). This protein is crystalized in a polymeric form, but only one chain (chain A) was used for simulations to reduce the computational cost. Despite being distal from the active site, the X-ray crystal structure mutations, P1K, Q258C, and E478Q mutations were modified to retain the native wild type structure. Co-crystalized solvent molecules were also deleted. Lamivudine was then superimposed against zidovudine (**Figure 7.2.**); the latter was then removed, leaving lamivudine sitting in the RT active site. Methionine

(Met) at position 184 was then manually mutated into isoleucine (Ile). The Chimera software package (https://www.cgl.ucsf.edu/chimera/) was used for ligand superimposition as well as amino acid residue mutation.
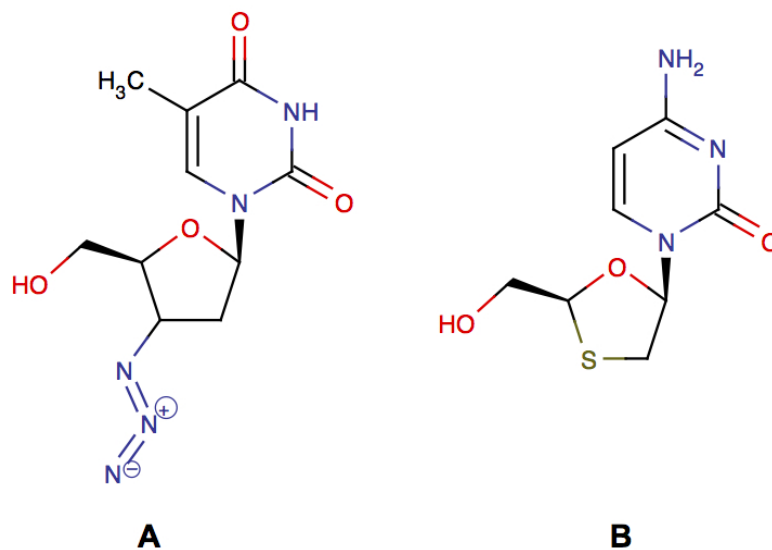
Figure 7.1. 2D structures of zidovudine and lamivudine (3TC), A and B, respectively (AZT).
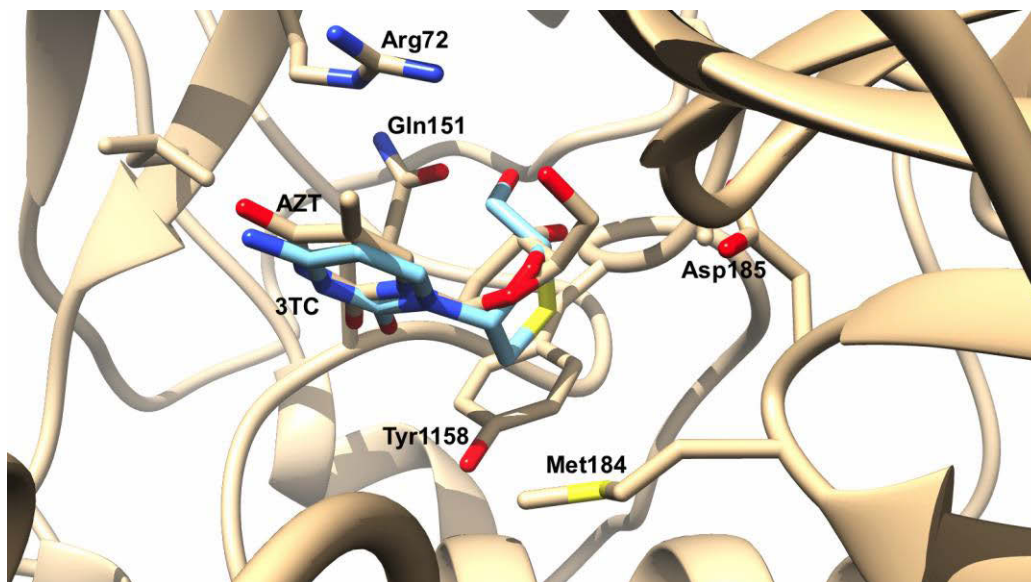
Figure 7.2. Lamivudine (3CT, cyan-colored carbon backbone) is superimposed against the Zidovudine (AZT, gold-colored carbon backbone) in the active site pocket of RT (PDB code: 1RTD)

7.3.2. Molecular Dynamics Simulation

A multiple MD-trajectory approach was adapted to understand the impact of M184I mutation on lamivudine binding with HIV-RT. A long MD run might lead to high statistical errors with denatured protein configuration during simulation period. In view of aforesaid, a multiple MD simulation acts as a high reliable alternative approach which lowers down force-field induces artifacts, statistical bias and computational time [50]. In this case a multiple MD approach was adapted for both the systems from an initial starting configuration generated from a 2ns MD run, followed by 4 different individual MD runs for a timescale of 5 ns with different initial velocities in each case. The multiple MD procedure adapted in this study was summarized in **Figure 7.3**.

```
        ┌─────────────────────────────┐
        │   Initial 2 ns MD simulation │
        └─────────────────────────────┘
                      ↓
        ┌─────────────────────────────┐
        │ 4 different starting configurations were │
        │     selected at 500 ps interval │
        └─────────────────────────────┘
          ↙        ↓        ↓        ↘
      (Conf.   (Conf.   (Conf.    (Conf.
        1)       2)       3)        4)
          │        │        │        │
        ┌─────────────────────────────┐
        │ 4 X 5 ns MD simulations starting from each configuration │
        │        with different velocities │
        └─────────────────────────────┘
                      ↓
        ┌─────────────────────────────┐
        │ Analysis of the individual trajectories as well as the │
        │                average │
        └─────────────────────────────┘
```
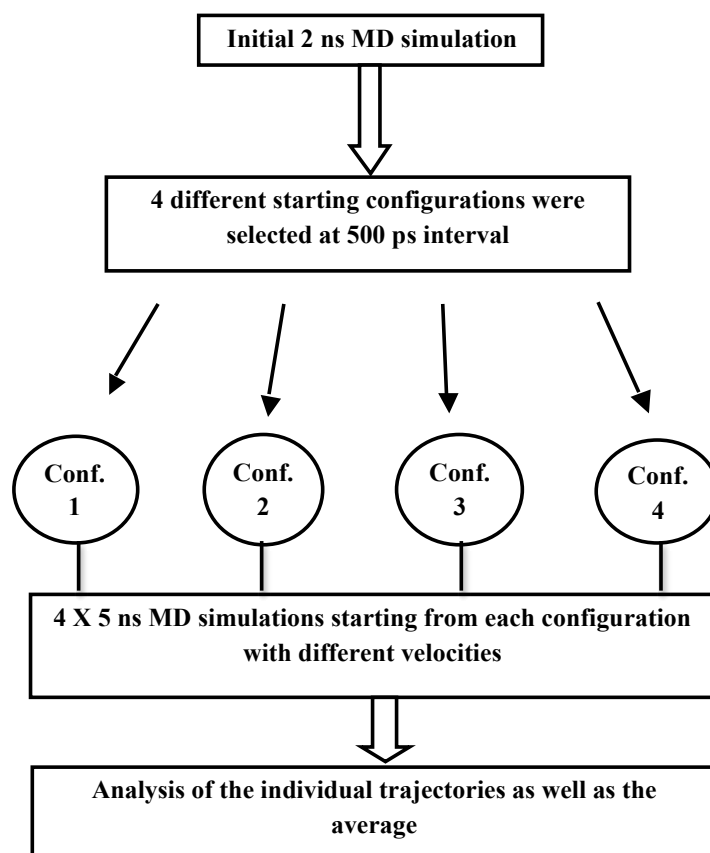
Figure 7.3. Multiple MD-trajectory approach adopted in this report

7.3.2.1. MD simulations set up and parameters

The MD simulations for the Lamivudine-RT complexes (i.e. wild-type and mutant) were performed using the GPU version of the PMEMD engine provided with the Amber 12 package [51-53]. Two protonation states ionized and unionized; of the triphosphate moieties attached to the ligand were considered (See Results and Discussion and Supplementary Material 2). The FF99SB variant of the AMBER force field [54] was used to describe the protein. The restrained electrostatic potential (RESP) procedure [55] was used to calculate the partial atomic charges for the ligands at the HF/6-31G* level using the Gaussian 03 package [56]. The LEAP module in Amber 12 was used for addition of hydrogen atoms to the protein and the addition of counter ions for neutralization. The systems were embedded within an orthorhombic box of TIP3P [57] water molecules such that no protein atom was within 8 Å of any box edge. Periodic boundary conditions were enforced and the long-range electrostatic interactions were treated with the particle-mesh Ewald method [58] incorporated in Amber12 with a direct space and vdW cut-off of 12 Å. The following MD procedure was applied to the initial 2 ns MD as well as the individual 5 ns MD runs (see Figure 3). Initial energy minimization of each system, with a restraint potential of 500 kcal/mol $Å^2$ applied to the solute, was performed for 1000 steps of steepest descent followed by 1000 steps of conjugate gradient minimization. A further 1000 steps of unrestrained conjugate gradient minimization of the entire system were then carried out. Canonical ensemble (NVT) MD simulations were then carried out for 50 ps, with gradual heating from 0 to 300K with harmonic restraints of 5 kcal/mol $Å^2$ applied to all solute atoms and a Langevin thermostat with a random collision frequency of 1/ps. The systems were subsequently equilibrated at 300 K in the NPT ensemble for 500 ps, during which no restrains were imposed and a Berendsen barostat was used to maintain the pressure at 1 bar. The SHAKE algorithm [59] was used to constrain the bonds of all hydrogen atoms , a time step of 2 fs and the SPFP precision model[60, 61] was used for all MD runs. A short production run for 2ns was performed followed by a multiple MD production run of 4*5ns from the last restart file (configuration generated from 2ns MD run) in an isothermal isobaric (NPT) ensemble using a Berendsen barostat for each case[62] with a target pressure of 1 bar and a pressure coupling constant of 2 ps. Coordinates were saved every 1 ps and the trajectories were analyzed every 1 ps using the PTRAJ module implemented in Amber 12.

7.3.3. Thermodynamic Calculations

The binding free energies of Lamivudine and RTs, wild type and mutant, were computed using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) approach [63-66]. For each 5 ns trajectory, the binding free energy was averaged over 1000 snapshots at 5 ps intervals. Average values over the 4 trajectories were also computed (see Table 1). The following set of equations describes the calculation of the binding free energy:

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

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

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

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

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

Where E_{gas} is the gas-phase energy; E_{int} is the internal energy; and E_{ele} and E_{vdW} are the Coulomb and Van der Waals energies, respectively. E_{gas} is evaluated directly from the ff03 force field terms. The solvation free energy, denoted by G_{sol}, can be decomposed into polar and nonpolar contribution states. The polar solvation contribution, G_{GB}, is determined by solving the GB equation, whereas, G_{SA}, the nonpolar solvation contribution is estimated from the solvent accessible surface area (SASA) determined using a water probe radius of 1.4 Å. T and S correspond to temperature and total solute entropy, respectively.

In order to determine the individual amino acid contribution towards total binding free energy between the Lamivudine inhibitor and wild-type/mutant HIV-RT types, a decomposition analysis of the interaction energy for each residue was computed by using the MM/GBSA binding free energy decomposition protocol in Amber 12.

7.3.4. Principle Component Analysis (PCA)

Before processing the MD trajectories for PCA, each 5 ns MD trajectories of the wild type and mutant complexes were stripped of solvent and ions using the PTRAJ module in AMBER 12.0. The stripped trajectories were then aligned against the fully minimized structure. PCA was performed for C-α atoms on 1000 snapshots each at 5 ps intervals. Using in-house scripts, the

first two principal components were calculated and the covariance matrices were generated. The first two principal components (PC1 and PC2) generated from each of the 4 individual 5 ns trajectories were averaged for both wild and M184I mutant. The first two principal components correspond to the first two Eigen vectors of the covariance matrix. The PCA scatter plots were then generated using the Xmgrace program (http://plasma-gate.weizmann.ac.il/Grace/). All structural diagrams were created using VMD [67]. The porcupine plots were generated using the ProDy interface of normal mode wizard (NMW) of VMD [68].

7.3.5. Residue Interaction Networks (RIN) analysis

The fully minimized structure of each system, wild and M184I mutant was used for constructing the RIN interactively in 2D graphs. The PROBE [36] software was used to identify the contacts (i.e. non-covalent interactions) between amino acids in the proteins by evaluating their atomic packing using small-probe contact dot surfaces. PROBE uses a small virtual probe (typically 0.25 Å) that is rolled around the van der Waals surface of each atom and an interaction (contact dot) is detected if the probe touches another non-covalently bonded atom [36].

7.3.5.1. Interactive visual analysis of residue networks

The RIN generated from the MD averaged structures were used to visualize the network using Cytoscape [69] and the plugin RINalyzer [46]. In a RIN, the nodes represent the protein amino acid residues and the edges between them represent the non-covalent interactions. The edges are labeled with an interaction type, usually including interatomic contact, hydrogen bonds, salt bridges etc.

7.4. Results and Discussion

7.4.1. MD simulations and systems stability

RMSD and potential energy fluctuations were monitored to ensure that the systems were well equilibrated before any further MD analysis. RMSD and potential energy plots are provided as Supplementary Material (**Supplementary Materials 1**).

All corresponding PDB coordinates of the simulated systems (wild and mutant for unionized and ionized phosphate moiety - 4 systems in total) are provided with the supplementary Material files.

7.4.2. Post-dynamic analysis: wild type versus M184I mutant

7.4.2.1. Root of mean square fluctuation (RMSF)

Figure 7.4 shows the residue based root mean square fluctuations (RMSF) of the wild type and M184I mutant simulations. The most interesting observation is that the amino acid residues in the region 170-220 (which contains the mutation site at position 184) show higher fluctuation in the mutant than compared with the wild type. One possible explanation is that the Ile184 residue in the mutant interacts less with the surrounding than methionine.

Interestingly, the M184I mutation was also found to impact the overall dynamics of some distal amino acid regions (**Figure 7.4.**). For instances, residues 2-50 exhibit larger fluctuations while residues 300-400 were found to show less fluctuation relative to the wild type.
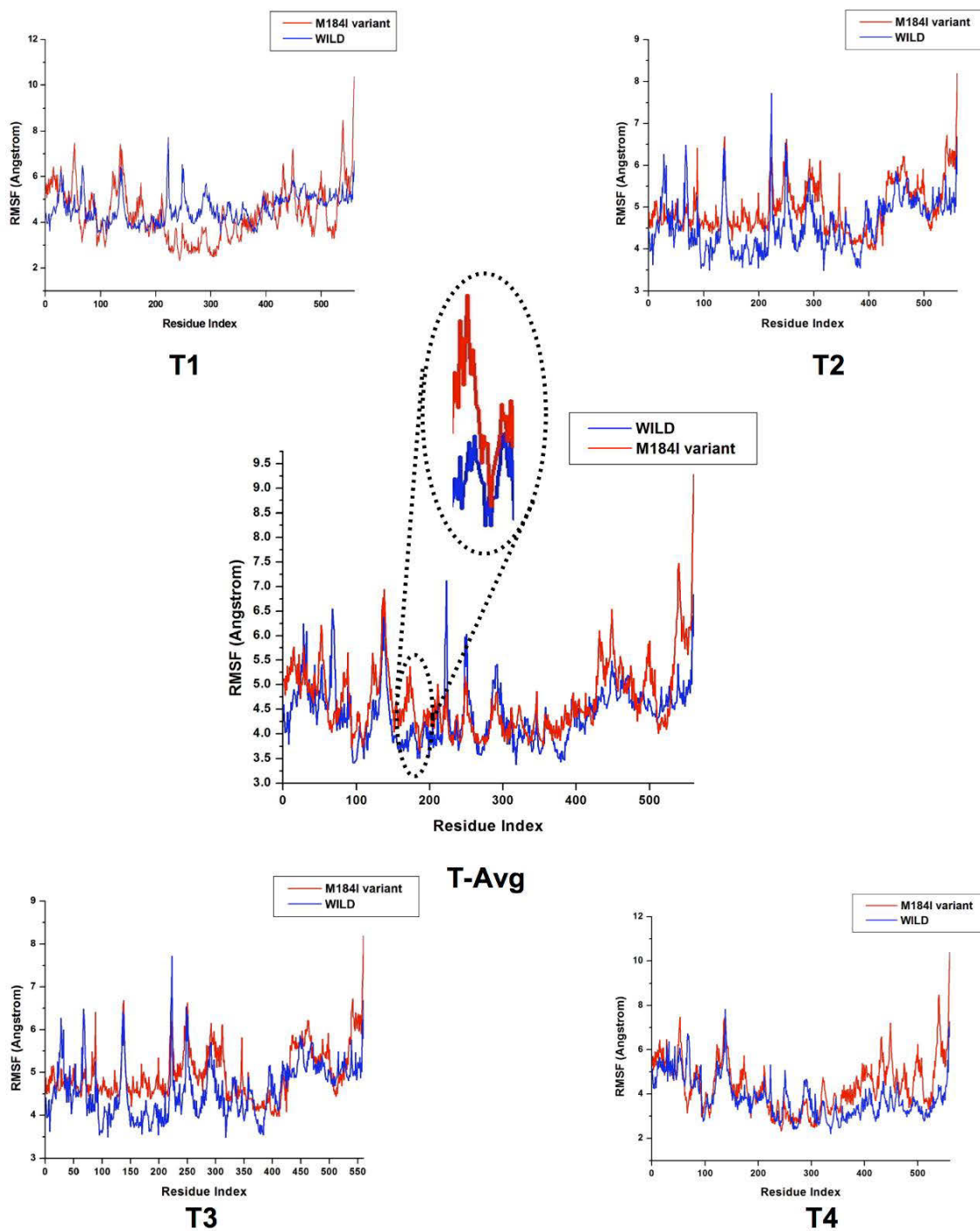
Figure 7.4. RMSF for the wild type and M184I mutant Lamivudine-RT complex systems: T1, T2, T3, T4 and T-Avg denote for the 4 individual 5 ns MD trajectories and the overall average, respectively. A zoomed view of region that contains the point mutation, 170-200, is shown in T-Avg plot.

7.4.2.2. Steric conflict between Ile184 and oxathiolane ring of lamivudine

Steric conflict between the oxathiolane ring of lamivudine and the side chain of beta-branched amino acids Ile at position 184 has traditionally been assumed to be responsible for the perturbed inhibitor binding. However, this assumption was based on static models [16]. In this study to test this assumption we monitored the distance between the (O) atom of oxathiolane ring and the C of the amino acid residue at position 184 (Met in wild type and Ile in mutant) during the MD trajectory. **Figure 7.5** suggests that substitution of Met with Ile results in increased steric conflict with the ligand that continues to exist during the duration of the MD simulation as evident from the increased separation between the (O) atom of oxathiolane ring and the C of the amino acid residue of Ile184. These findings lend credence to the static based assumption of Sarafianos *et al* [16].
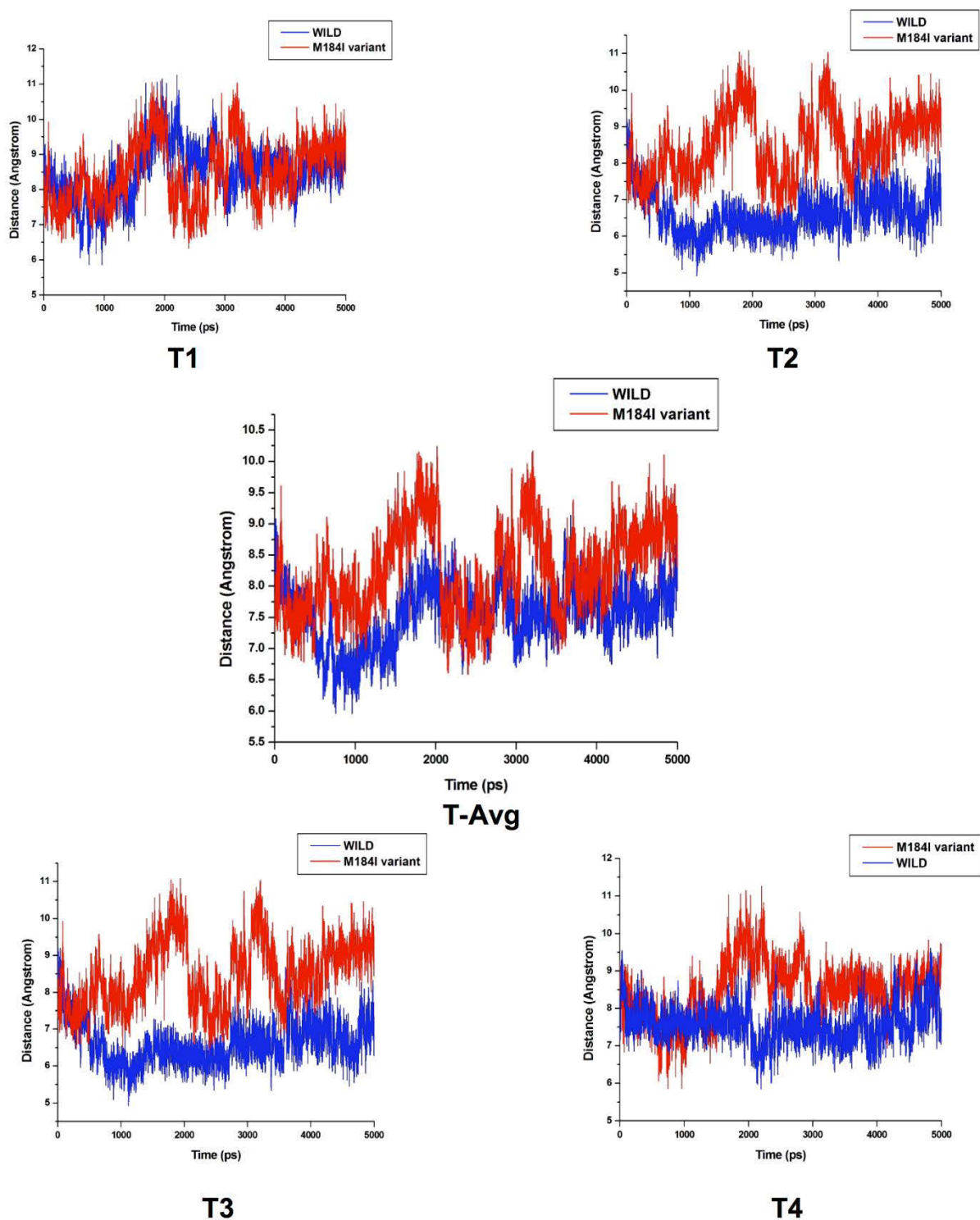
Figure 7.5. The distance between the (O) atom of oxathiolane ring and the C☐ of the amino acid residue at position 184 (Met or Ile). T1, T2, T3, T4 and T-Avg denote for the 4 individual 5 ns MD trajectories and the overall average, respectively.

7.4.3. MM/PBSA binding free energy calculations

All the components of molecular mechanics and solvation energy were averaged over a 4*5ns multiple MD simulation in each case using the MM/GBSA technique are listed in **Table 7.1**. The calculated binding free energy (ΔG_{bind}) between lamivudine and wild-type RT is -45.6597 kcal/mol compared to -38.5251 kcal/mol in the case of M184I mutant. Such a large reduction in binding affinity (~8 kcal/mol) as a result of mutation could significantly impair ligand binding and thus the effectiveness of lamivudine against the mutant – this is in a great accordance with experimental data (the EC_{50} value in case of M184I mutant is ~1000 times lower than the wild type) [70]. The calculated van der Waals contributions (ΔE_{vdW}) to the binding free energy in the lamivudine wild-type bound RT complex (-48.2095 kcal/mol) are higher than that for the lamivudine bound M184I mutant RT complex (-42.5814 kcal/mol). On the other hand, the calculated electrostatic contributions (ΔE_{ele}) to the binding free energy for lamivudine bound M184I mutant RT complex (-16.9961 kcal/mol) are lower compared to that for the lamivudine wild-type bound RT complex (-24.6420 kcal/mol). In the lamivudine wild-type bound RT complex, the calculated solvation contributions ($\Delta G_{sol.}$ $\Delta G_{sol} = \Delta G_{SA} + \Delta G_{GB}$) to the binding free energy (26.4379 kcal/mol) are higher than that for the lamivudine bound M184I mutant RT complex (21.0511 kcal/mol). The free energy components shown in **Table 7.1** suggests that the majority of the favorable contributions observed for lamivudine binding arise from ΔE_{vdW} and $\Delta E_{ele.}$ The remarkable difference (~13 kcal/mol) in the ΔG_{gas} as a result of mutation also confirmed that M184I mutation could lead to loss of efficacy of lamivudine.

Table 7.1. The calculated binding free energies based on MM/GBSA method. T1, T2, T3, T4 and T-Avg denote for the 4 individual 5 ns MD trajectories and the overall average, respectively.

Complexes	Trajectory	ΔG_{bind}	ΔE_{ele}	ΔE_{vdW}	ΔG_{gas}	ΔG_{sol}
Wild	T1	-46.5489±0.1251	-25.0332±0.1442	-48.3440±0.1012	-73.3772±0.1851	26.8283±0.1104
	T2	-44.7706±0.1838	-24.2509±0.2109	-48.0673±0.1393	-72.3182±0.2853	27.5476±0.1630

	T3	-44.7706±0.1838	-24.2509±0.2109	-48.0673±0.1393	-72.3182±0.2853	27.5476±0.1630
	T4	-46.5489±0.1251	-25.0332±0.1442	-48.3440±0.1012	-73.3772±0.1851	26.8283±0.1104
	Tavg	**-45.6597±0.1544**	**-24.6420±0.1775**	**-48.2095±0.1202**	**-72.8477±0.2352**	**26.4379±0.1367**
Mutant	T1	-38.2631±0.1157	-16.2502±0.2371	-42.1752±0.1006	-58.4254±0.2612	20.1624±0.2386
	T2	-38.2631±0.1157	-16.2502±0.2371	-42.1752±0.1006	-58.4254±0.2612	20.1624±0.2386
	T3	-38.2631±0.1157	-16.2502±0.2371	-42.1752±0.1006	-58.4254±0.2612	20.1624±0.2386
	T4	-39.3164±0.1290	-19.2339±0.2360	-43.8000±0.1177	-63.0339±0.2549	23.7175±0.2088
	Tavg	**-38.5251±0.1190**	**-16.9961±0.2368**	**-42.5814±0.1049**	**-59.5775±0.2569**	**21.0511±0.2311**

For the un-protonated triphosphate moiety, interestingly, however, the same trend was observed in the overall binding affinity, the electrostatic contribution was higher when compared to the protonated state – this might be due to higher electrostatic interactions between the active site residues with the charged triphosphate moiety (see **Supplementary Material 2**).

7.4.3. Validation of the binding energy calculations

To validate the binding free energy results, another RT mutation, M184V, with available experimental data[71] was considered. Encouragingly, the calculated binding free energy was

found to be in a good agreement with experimental data (see **Supplementary Material**). Such findings imply that the simulation protocol of this work is appropriate and reliable.

7.4.4. Per-residue interaction energy decomposition analysis

The binding free energy was further decomposed into contributions from each HIV-RT amino acid residue. In **Table 7.2**, the comparison of protein-ligand interaction spectra between wild-type and M184I mutant are shown. It can be observed from the energy decomposition analysis that in the lamivudine bound wild-type HIV-RT system, the major contributions were -3.21, -2.16 and -1.14 kcal/mol from amino acid residues Arg72, Tyr115, and Met184, respectively (**Table 7.2.**). On the other hand, there were some minor contributions towards the interaction energy from residues Asp110 (-0.13 kcal/mol), Ala114 (-0.79 kcal/mol) and Phe116 (-0.83 kcal/mol).

As shown in **Table 7.2** below, the decomposed van der Waals and electrostatic energies for residues 72, 115 and 184 in the lamivudine bound M184I mutant complex implies that the change in the van der Waals contribution is what is mostly responsible for the decrease in the interaction energy. Furthermore, when Met is mutated to Ile at position 184, the prominent van der Waals contributions towards the total binding free energy from this residue increased by two-fold with a slight decrease in electrostatic contribution but it ultimately lowers the total van der Waals contributions of other prominet residues which ultimately affects the total binding free energy. This finding stands consistent with decreased van der Waals contributions from Arg72 (-3.21 to -2.07 kcal/mol) and Tyr115 (-2.16 to -2.10 kcal/mol) and an overall loss of van der Waals contribution towards binding free energy (**Table 7.1. and 7.2.**).

Table 7.2. The decomposed van der Waals and electrostatic energies (kcal/mol) for residues 72, 113, 115 and 184 in Lamivudine bound wild-type and M184I mutant complexes (values were averaged over the 4 individual trajectories).

Residues	van-der Waals	Electrostatic
Arg72	-3.21±0.28	-2.01±1.53
	*-2.07±0.42**	*-0.18±1.33**
Met184	-1.14±0.21	0.16±0.17
Ile184	*-2.39±0.52**	*0.07±0.37**
Tyr115	-2.16±0.18	-0.19±0.14
	*-2.10±0.42**	*-0.32±0.30**

mutant

Interestingly, as evident from **Table 7.2**, mutation of Met to Ile at position 184 has improved the binding at the site of mutation - this might be due to better hydrophobic interactions with the Ile side chain – the overall drug binding affinity is reduced due to the negative impact of the mutation on the binding affinity of the nearby active residues. Different binding forces between the ligand and the protein, wild and mutant are shown in **Figure 7.6**.
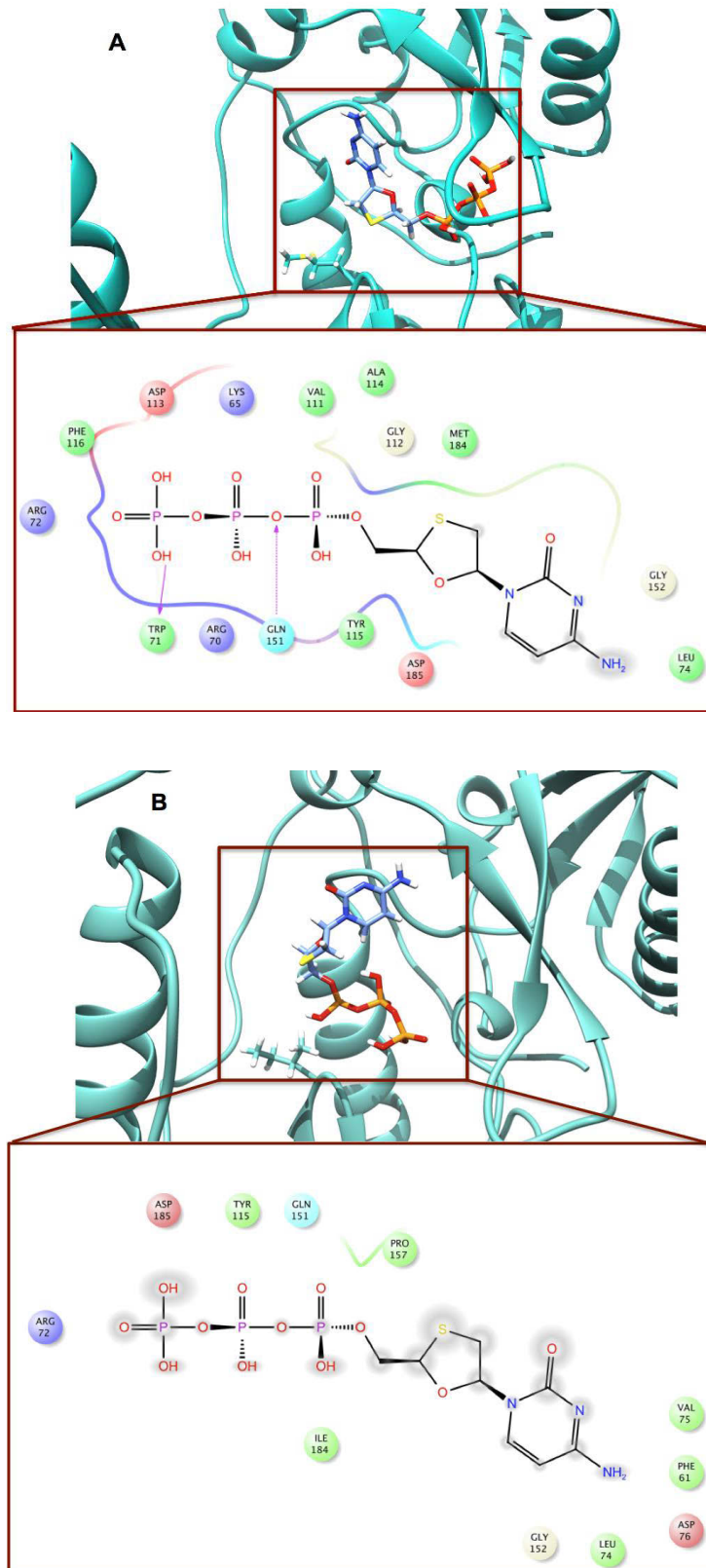
Figure 7.6. Representative structures for the lamivudine-RT complexes: wild type, A and M184I, B, respectively with graphical representation of the different binding forces.

7.4.5. Principle component analysis (PCA)

Figure 7.7, shows a PCA scatter plot generated for the wild type and M184I mutant showing a significant difference between both systems as evident from the characteristic structures plotted along the direction of two principal components. From the scatter plot, it is clear that eigenvectors computed from the MD trajectory for both systems are quite varied which clearly indicates difference in protein motion between wild and M184I mutant.
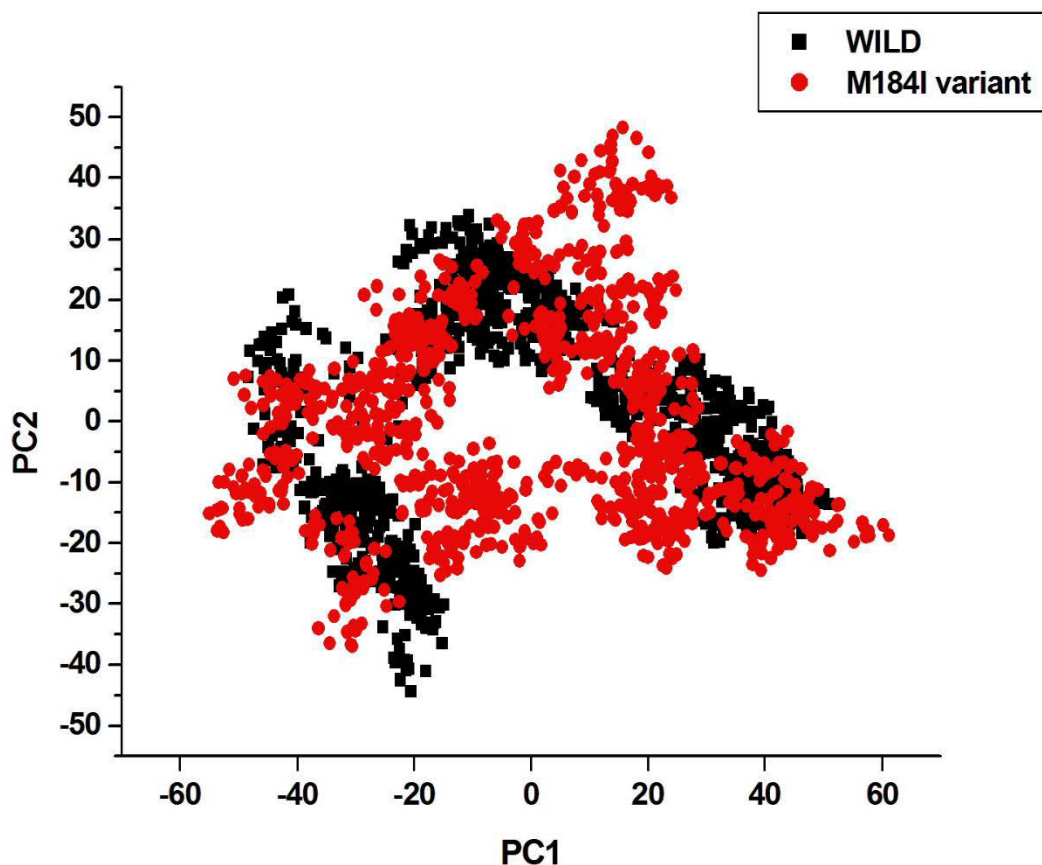
Figure 7.7. PCA scatter plot of 1000 snapshots along the pair of first two principal components, PC1 and PC2 for wild type and M184I mutant showing difference in eigenvectors. Eigen values were averaged over the 4 individual 5 ns MD trajectories.

In most cases a few low frequency normal modes are enough to capture major protein motions along specific directions, which are represented by eigenvectors. In this case we generate porcupine plots with three low frequency modes (Mode=1, 2 & 3) to visualize the comparative motion difference between wild type and mutant (**Figure 7.8.**). The eigenvectors show a clear difference in direction of motion, which is consistent with the PCA, scatter plot.
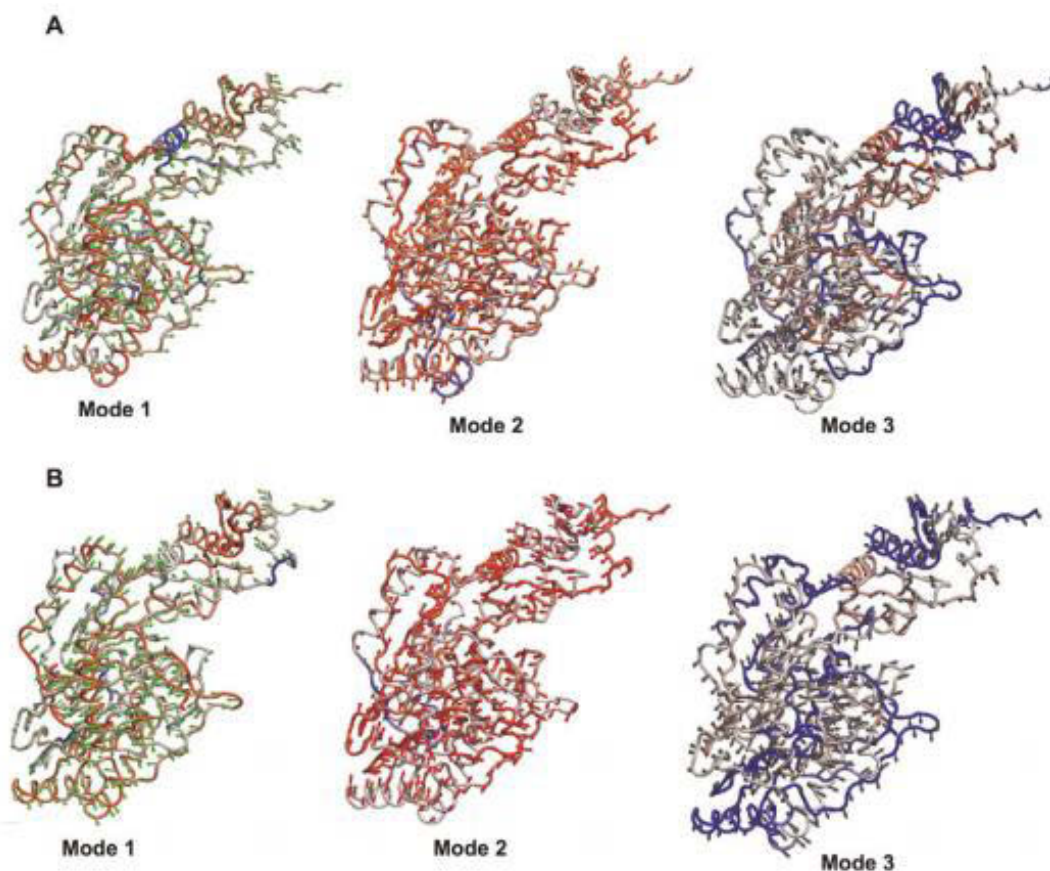
Figure 7.8. Porcupine plots showing atomic components for wild, A and M184I mutant, B, in different active modes. The green, red, grey arrows represent eigenvectors showing direction of motions across Mode 1; Mode 2 & Mode 3, respectively.

7.4.6. Residue interaction network (RIN)

The network analysis of the protein backbone is a new strategy to identify key residue interactions and can be used to explore the difference in residue interaction networks between different proteins including wild type and mutants. In this work we investigated the relationship between key residues of the wild type and M184I mutant by generating residue interaction

networks using the representative structures from the 5ns of MD. As evident from the RIN plots (**Figures 7.9. and 7.10.**), it is clear that the M184I mutation has distorted the overall residue interaction network when compared to wild type. For instance, as shown in Figures 7 and 8, there is a hydrogen bond interaction between Tyr115 and Met184 whereas in the case of the mutant where Met184 has been mutated to Ile184 the hydrogen bond interaction with Tyr115 changes to just a close atom interaction. It is interesting to observe that in the case of the wild type, Pro157 builds a close atom interaction with Met184 whereas in the case of the mutant there is a close atom interaction instead between Gln161 and Ile184. Again, the M184I mutation affects the interaction network, which ultimately affects the protein backbone and consequently the drug binding landscape.

Figure 7.9. Residue Interaction Network showing close atom interactions between Ile184 and Tyr115 in case of M184I mutant, A, and hydrogen bond interaction between Met184 and Tyr115 in case of wild type RT, B.

Figure 7.10. Residue Interaction Network showing close atom interactions between Gln161 & Ile184 in case of M184I mutant, C and close atom interaction between Pro157 & Met184 in case of wild type RT, D.

7.5. Conclusions

The precise molecular understanding of the serious impact of a single M184I on the RT resistance to lamivudine is lacking in literature. In this report, we embarked on a wide range of computational approaches in order to provide a multi-dimensional view on M184I resistance to lamivudine. Multiple molecular dyn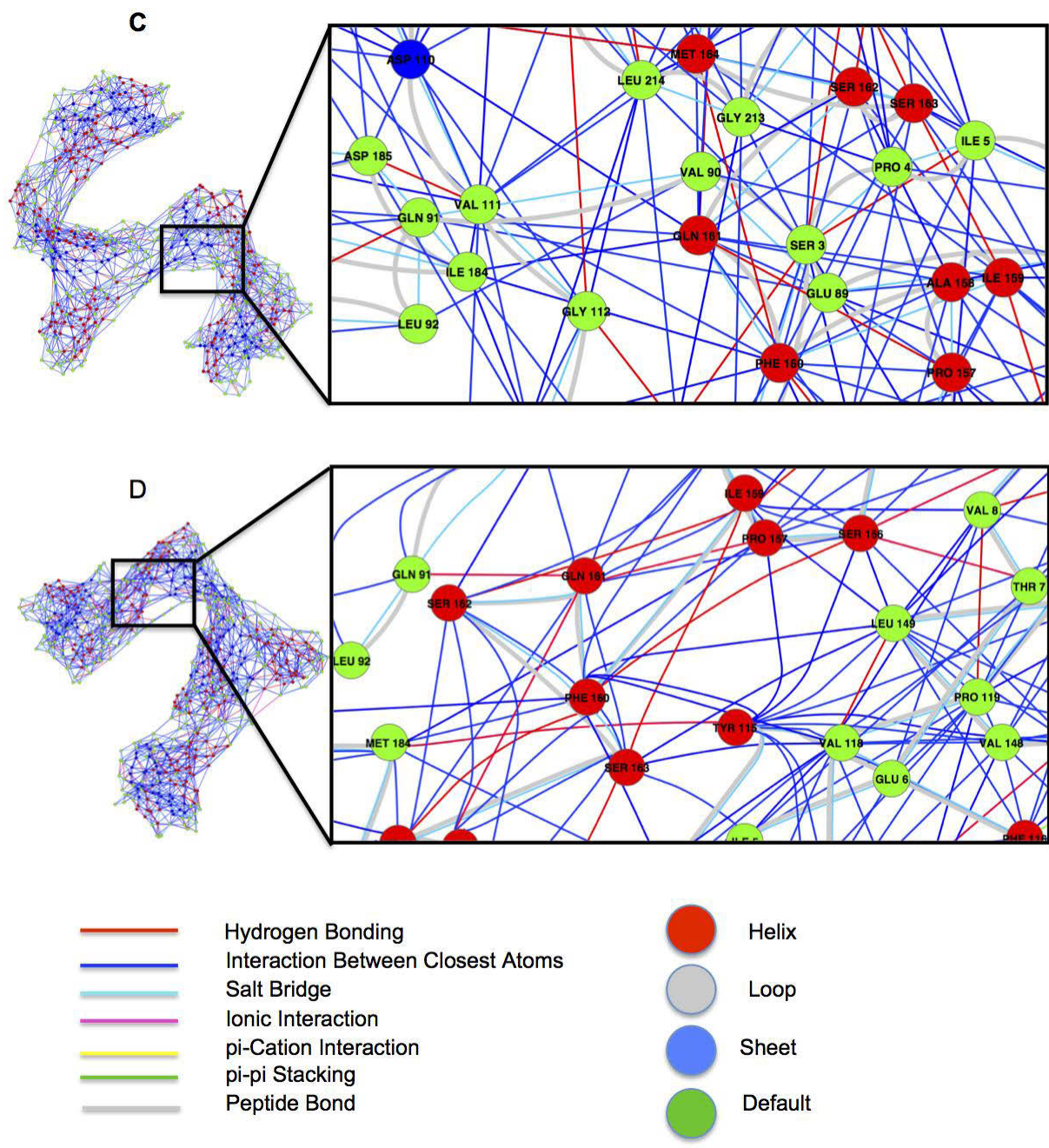amics simulations, binding free energy calculations, principle component analysis (PCA) and residue interaction network (RIN) analysis led us to several findings that can explain the serious impact of M184I mutation on drug resistance. These findings verified that mutation decreased drug binding affinity by ~8 kcal/mol, distorted the ligand optimum orientation in the RT active site, affected the overall enzyme conformational landscape and distorted the atomic interaction network with the ligand.

The findings of this report can provide potential markers for further design of novel inhibitors that are less susceptible to drug resistance.

7.6. Supplementary Materials

RMSD vs. time and Potential Energy (kcal/mol) vs. time data for ionized and unionized form of wild and M184I variant of HIV-RT complexed with lamivudine as well as of M184V variant is provided with Supplementary Materials 1 and Supplementary Materials 2.

7.7. Acknowledgements

7.8. Conflicts of Interest

Authors declare no potential conflicts of interest.

7.9. References

1. UNAIDS. (2011) World AIDS day report 2011, *http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/jc2216_worldaidsday_report_2011_en.pdf*, 11 Nov. 2013.

2. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.

3. 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 & Drug Design 81*, 208-218.

4. Morah, E. U. (2007) Are People Aware of Their HIV-positive Status Responsible for Driving the Epidemic in SubSaharan Africa? The Case of Malawi In *Development Policy Review*, pp 215-242.

5. Pani A, L., A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. . (2002) Targeting HIV: Old and New Players, *Current Drug Target -Infectious Disorders 2*, 17-32.

6. Esposito, F., Corona, A., and Tramontano, E. (2012) HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions, *Molecular Biology International 2012*, 1-23.

7. Schinazi, R. F., Hernandez-Santiago, B. I., and Hurwitz, S. J. (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses (vol 71, pg 322, 2006), *Antiviral Research 72*, 256-256.

8. Ilina, T., LaBarge, K., Sarafianos, S. G., Ishima, R., and Parniak, M. A. (2012) Inhibitors of HIV-1 Reverse Transcriptase—Associated Ribonuclease H Activity, *Biology 1*, 521-541.

9. Vivet-Boudou, V., Didierjean, J., Isel, C., and Marquet, R. (2006) Nucleoside and nucleotide inhibitors of HIV-1 replication, *Cellular and Molecular Life Sciences 63*, 163-186.

10. Bauman, J. D., Das, K., Ho, W. C., Baweja, M., Himmel, D. M., Clark, A. D., Jr., Oren, D. A., Boyer, P. L., Hughes, S. H., Shatkin, A. J., and Arnold, E. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design, *Nucleic Acids Research 36*, 5083-5092.

11. Turner, D., Brenner, B., and Wainberg, M. A. (2003) Multiple effects of the M184V resistance mutation in the reverse transcriptase of human immunodeficiency virus type 1, *Clinical and Diagnostic Laboratory Immunology 10*, 979-981.

12. Ray, A. S. Intracellular interactions between nucleos(t)ide inhibitors of HIV reverse, *D - 101134876*, - 113-125.

13. Hamers, R. L., Kityo, C., Sigaloff, K. C., and de Wit, T. F. R. (2013) Pretreatment HIV-1 drug resistance in Africa, *The Lancet infectious diseases 13*, 476-476.

14. Wainberg, M. A., and Turner, D. (2004) Resistance issues with new nucleoside/nucleotide backbone options, *Jaids-Journal of Acquired Immune Deficiency Syndromes 37*, S36-S43.

15. Gao, H. Q., Boyer, P. L., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2000) The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase, *Journal of Molecular Biology 300*, 403-418.

16. Sarafianos, S. G., Das, K., Clark, A. D., Ding, J. P., Boyer, P. L., Hughes, S. H., and Arnold, E. (1999) Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids, *Proceedings of the National Academy of Sciences of the United States of America 96*, 10027-10032.

17. Bahareh, H., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2013) Integrated Approach to Structure-Based Enzymatic Drug Design: Molecular Modeling, Spectroscopy, and Experimental Bioactivity, *Chemical Reviews DOI: 10.1021/cr300314q*.

18. Aruksakunwong, O., Wolschann, P., Hannongbua, S., and Sompornpisut, P. (2006) Molecular dynamic and free energy studies of primary resistance mutations in HIV-1 protease-ritonavir complexes, *Journal of Chemical Information and Modeling 46*, 2085-2092.

19. Hou, T. J., 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.

20. 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.

21. Zhou, Z. G., Madrid, M., Evanseck, J. D., and Madura, J. D. (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase, *Journal of the American Chemical Society 127*, 17253-17260.

22. Xue, W. W., Qi, J., Yang, Y., Jin, X. J., Liu, H. X., and Yao, X. J. (2012) Understanding the effect of drug-resistant mutations of HIV-1 intasome on raltegravir action through molecular modeling study, *Molecular Biosystems 8*, 2135-2144.

23. Chachra, R., and Rizzo, R. C. (2008) Origins of resistance conferred by the R292K neuraminidase mutation via molecular dynamics and free energy calculations, *Journal of Chemical Theory and Computation 4*, 1526-1540.

24. Liu, H. X., Yao, X. J., Wang, C. Q., and Han, J. A. (2010) In Silico Identification of the Potential Drug Resistance Sites over 2009 Influenza A (H1N1) Virus Neuraminidase, *Molecular Pharmaceutics 7*, 894-904.

25. Guo, Z. Y., Prongay, A., Tong, X., Fischmann, T., Bogen, S., Velazquez, F., Venkatraman, S., Njoroge, F. G., and Madison, V. (2006) Computational study of the effects of mutations A156T, D168V, and D168Q on the binding of HCV protease inhibitors, *Journal of Chemical Theory and Computation 2*, 1657-1663.

26. Pan, D. B., Xue, W. W., Zhang, W. Q., Liu, H. X., and Yao, X. J. (2012) Understanding the drug resistance mechanism of hepatitis C virus NS3/4A to ITMN-191 due to R155K, A156V, D168A/E mutations: A computational study, *Biochimica Et Biophysica Acta-General Subjects 1820*, 1526-1534.

27. Cheng, X. L., Cui, G. L., Hornak, V., and Sinnnerling, C. (2005) Modified replica exchange simulation methods for local structure refinement, *Journal of Physical Chemistry B 109*, 8220-8230.

28. Affentranger, R., Tavernelli, I., and Di Iorio, E. E. (2006) A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling, *Journal of Chemical Theory and Computation 2*, 217-228.

29. Okur, A., Wickstrom, L., Layten, M., Geney, R., Song, K., Hornak, V., and Simmerling, C. (2006) Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model, *Journal of Chemical Theory and Computation 2*, 420-433.

30. Liu, P., Kim, B., Friesner, R. A., and Berne, B. J. (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water, *Proceedings of the National Academy of Sciences of the United States of America 102*, 13749-13754.

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

32. Thomas, J. R., Gedeon, P. C., Grant, B. J., and Madura, J. D. (2012) LeuT Conformational Sampling Utilizing Accelerated Molecular Dynamics and Principal Component Analysis, *Biophysical Journal 103*, L01-L03.

33. Amadei, A., Linssen, A. B., de Groot, B. L., van Aalten, D. M., and Berendsen, H. J. (1996) An efficient method for sampling the essential subspace of proteins, *Journal of Biomolecular Structure and Dynamics 13*, 615-625.

34. van Aalten, D. M., Findlay, J. B., Amadei, A., and Berendsen, H. J. (1995) Essential dynamics of the cellular retinol-binding protein– evidence for ligand-induced conformational changes, *Protein Engineering, Design and Selection 8*, 1129-1135.

35. Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins, *Proteins: Structure, Function, and Bioinformatics 17*, 412-425.

36. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms, *Journal of Molecular Biology 285*, 1711-1733.

37. Case, D. A. (1994) Normal-mode analysis of protein dynamics, *Current Opinion in Structural Biology 4*, 285-290.

38. Brooks, B., and Karplus, M. (1985) Normal-modes for specific motions of macromolecules - Application to the hinge-bending mode of lysozyme, *Proceedings of the National Academy of Sciences of the United States of America 82*, 4995-4999.

39. Laine, E., de Beauchene, I. C., Perahia, D., Auclair, C., and Tchertanov, L. (2011) Mutation D816V Alters the Internal Structure and Dynamics of c-KIT Receptor Cytoplasmic Region: Implications for Dimerization and Activation Mechanisms, *Plos Computational Biology 7*.

40. Teodoro, M. L., Phillips, G. N., and Kavraki, L. E. (2003) Understanding protein flexibility through dimensionality reduction, *Journal of Computational Biology 10*, 617-634.

41. Yang, L., Song, G., Carriquiry, A., and Jernigan, R. L. (2008) Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes, *Structure 16*, 321-330.

42. del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins, *Molecular Systems Biology 2*.

43. Welsch, C., Schweizer, S., Shimakami, T., Domingues, F. S., Kim, S., Lemon, S. M., and Antes, I. (2012) Ketoamide Resistance and Hepatitis C Virus Fitness in Val55 Variants of the NS3 Serine Protease, *Antimicrobial Agents and Chemotherapy 56*, 1907-1915.

44. Welsch, C., Domingues, F. S., Susser, S., Antes, I., Hartmann, C., Mayr, G., Schlicker, A., Sarrazin, C., Albrecht, M., Zeuzem, S., and Lengauer, T. (2008) Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus, *Genome Biology 9*.

45. Xue, W. W., Jin, X. J., Ning, L. L., Wang, M. X., Liu, H. X., and Yao, X. J. (2013) Exploring the Molecular Mechanism of Cross-Resistance to HIV-1 Integrase Strand Transfer Inhibitors by Molecular Dynamics Simulation and Residue Interaction Network Analysis, *Journal of Chemical Information and Modeling 53*, 210-222.

46. Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011) Analyzing and visualizing residue networks of protein structures, *Trends in Biochemical Sciences 36*, 179-182.

47. Doncheva, N. T., Assenov, Y., Domingues, F. S., and Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures, *Nature Protocols 7*, 670-685.

48. 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, *Chemical Biology & Drug Design 81*, 208-218.

49. Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Drug Development Research 74*, 283-295.

50. Kanibolotsky, D. S., Novosyl'na, O. V., Abbott, C. M., Negrutskii, B. S., and El'skaya, A. V. (2008) Multiple molecular dynamics simulation of the isoforms of human translation elongation factor 1A reveals reversible fluctuations between "open" and "closed" conformations and suggests specific for eEF1A1 affinity for Ca(2+)-calmodulin, *Bmc Structural Biology 8*.

51. Goetz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, *Journal of Chemical Theory and Computation 8*, 1542-1555.

52. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2012), AMBER 12, University of California, San Francisco.

53. Salomon-Ferrer, R., Goetz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, *Journal of Chemical Theory and Computation 9*, 3878-3888.

54. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins-Structure Function and Bioinformatics 78*, 1950-1958.

55. 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, *Journal of Computational Chemistry 16*, 1357-1377.

56. 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*.

57. 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.

58. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *Journal of Chemical Physics 103*, 8577-8593.

59. 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, *Journal of Computational Physics 23*, 327–341.

60. Le Grand, S., Götz, A. W., and Walker, R. C. (2013) SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations, *Computer Physics Communications 184*, 374-380.

61. Le Grand, S., Goetz, A. W., and Walker, R. C. (2013) SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations, *Computer Physics Communications 184*, 374-380.

62. 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, *Journal of Chemical Physics 81*, 3684–3690.

63. 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.

64. 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.

65. Tsui, V., and Case, D. A. (2000) Theory and applications of the Generalized Born solvation model in macromolecular simulations, *Biopolymers 56*, 275-291.

66. 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.

67. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics & Modelling 14*, 33-38.

68. Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: Protein Dynamics Inferred from Theory and Experiments, *Bioinformatics 27*, 1575-1577.

69. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Research 13*, 2498-2504.

70. Choo, H., Chong, Y., and Chu, C. K. (2003) The role of 2′,3′-unsaturation on the antiviral activity of anti-HIV nucleosides against 3TC-resistant mutant (M184V), *Bioorganic & Medicinal Chemistry Letters 13*, 1993-1996.

71. Diallo, K., Gotte, M., and Wainberg, M. A. (2003) Molecular impact of the M184V mutation in human immunodeficiency virus type 1 reverse transcriptase, *Antimicrobial Agents and Chemotherapy 47*, 3377-3383.

CHAPTER 8

Identification, binding mode and prospective chemical structural features of novel Nef protein inhibitors as potential anti-HIV drugs
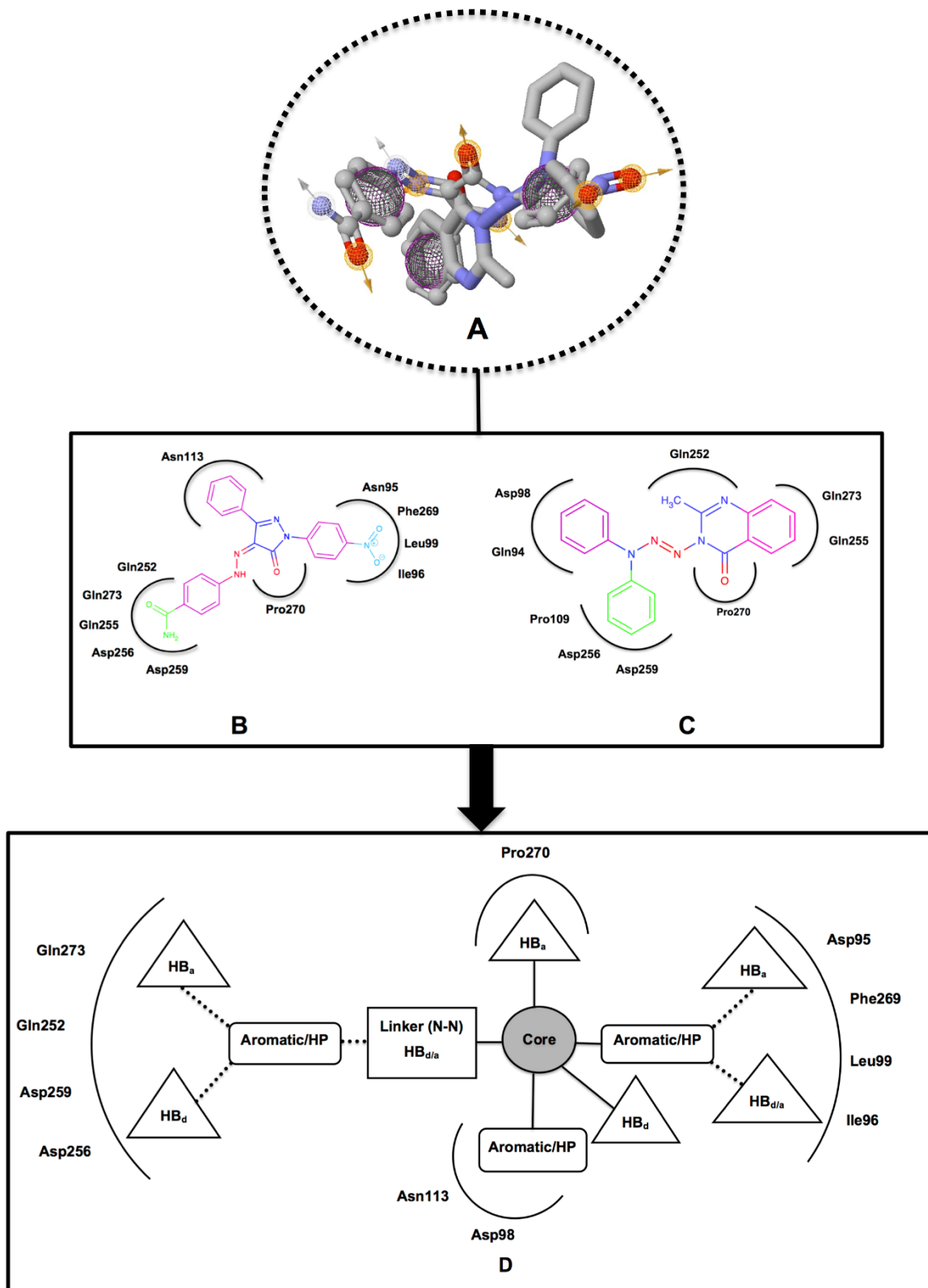
Suri Moonsamy[a], Soumendranath Bhakat[a] and Mahmoud E. S. Soliman[a*]

[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

8.1. Graphical Abstract

8.2. Abstract

As an emerging target, HIV-Nef protein, an accessory pathogenic factor, plays a significant role in HIV replication. Reports indicated that removal of Nef protein leads to high level of viral insensitivity. Therefore, targeting Nef is considered a key strategy towards HIV/AIDS treatment.

Up-to-date, only one compound has been reported as a potent Nef inhibitor, known as B9. This has prompted us to provide this first account of an integrated computational framework in order to identify more potential Nef inhibitors. Herein, using a hybrid ligand virtual screening, shape similarity- and pharmacophore-based, approach combined with molecular dynamics and post dynamics analysis were applied to identify potential new leads targeting HIV-Nef with a detailed analysis of their binding modes. The top ranked compounds from the shape similarity-based library (**ZINC04177596,** $\Delta G_{bind=}$ -28.7482 kcal/mol**)** and pharmacophore-based library (**ZINC36617540,** $\Delta G_{bind=}$ -20.2271 kcal/mol**)** possess comparatively better binding affinities than the reference molecule, B9 (ΔG_{bind}= -18.0694 kcal/mol). Both top identified hits, **ZINC04177596** and **ZINC36617540** showed similar binding mode at the active site as the prototype, B9. Hydrophobic and electrostatic interactions seemed to be the prominent binding forces that hold these ligands at the dimer interface of Nef protein. A set of chemical structural features that can be used as a guide in the design of potential Nef inhibitors is also highlighted herein.

We believe that information gained from this study would be of great importance in the discovery and design of potential small molecules targeting HIV-Nef.

Key words: Hybrid ligand-based virtual screening; HIV-Nef; HIV/AIDS; molecular dynamics (MD); binding free energy

Running title: HIV-Nef inhibitors

8.3. Introduction

Ever since the Acquired Immunodeficiency Syndrome (AIDS) was first reported in 1983, this global endemic has already proclaimed the lives of estimated 22 million individuals worldwide [1]. The Human Immunodeficiency Virus type 1 (HIV-1) responsible for causing AIDS still remains one of the most challenging classified infectious diseases [1,2]. Globally, it has been reported that an estimated 34 million people live with HIV/AIDS [3,4], and this includes approximately 22.9 million people that characterize this overall global estimate [3,4].

Despite the substantial effort made in HIV/AIDS research, no cure still exists for this global endemic. At present, the only effective therapeutic regimen comprises of multiple drug combinations that target different HIV enzymes at various stages of its life cycle [5]. These include reverse transcriptase (RT) inhibitors, protease inhibitors (PIs), integrase (IN) inhibitors, entry inhibitors and progression inhibitors [5,6].

Nef (Negative Factor) is an accessory gene product of HIV, which has an imperative role in viral replication [7,8]. It was shown that a large deletion in the Nef gene greatly reduces the severity of HIV [9]. The Nef is a small protein (~ 27-35 kDa) and myristoylated in nature [8]. Its primary mode of transport is localization from the cytosol into the host plasma membrane [8,9]. Nef plays a crucial role by down regulating the surface expression of CD4, MHC-I, MHC-II, and CD28, which are critical in the formation of immune synapses [10,11].

Recent drug discovery attempted to report the Nef protein as a crucial target in designing novel antiretroviral drugs. To date, only two previous studies have reported the concept of designing potential Nef inhibitors against HIV/AIDS [12,13]. These were carried out mainly using high-throughput screening (HTS) as well as some small-scale structure-based drug design approaches. One such example is the compound B9, which has been discovered from HTS appeared very effective at inhibiting the dimerization of the two Nef dimers and hence blocking Nef functionality. The binding pocket of B9 with Nef dimer has been identified as potential hot spot to design novel small molecule inhibitors [14,15]. Guanidine alkaloid-p53 interactions and other compounds interactions with HIV-Nef and SH3 domains have been classified as potential drug candidates in recent small-scale structure-based drug design techniques [5,13].

In recent years, molecular modeling techniques have become close counterparts between computation and experimental in biological contexts [16-21]. Numerous computational tools are now accessible for searching chemical databases for novel compounds with desired features. Among these tools is virtual screening, which has become an imperative tool in current drug discovery scenarios [22-25]. Virtual screening can be divided into ligand-based and structure-based approaches [25-27] (*19-22*). Ligand-based virtual screening methods use the knowledge of a known ligand and an unknown receptor site to estimate the binding affinities of potential hit leads for a given target [25-27]. However, structure-based virtual screening (SBVS) techniques use the knowledge of the 3D structure of a target protein in order to provide estimation of the binding affinities of small compounds towards the target of interest [28-31]. Even though, both techniques implement the use of different molecular knowledge, these approaches are complimentary in nature.

As indicated above, only one compound, known as B9, was reported to inhibit Nef dimerization and identified as a potential starting point for further optimization. To this end, the aim of this work is to identify more potential Nef inhibitors by exploiting the structural features of B9 using a framework of computational tools. These tools include a hybrid shape similarity and pharmacophore-based virtual screening approach, molecular dynamics simulations, post-dynamics analysis and binding free energy calculations.

It is worth mentioning that most virtual screening engines rely primarily on docking calculations. It is also widely known that results based solely on docking calculations, in many cases, could be artifacts and/or questionable[32]. Therefore, in this work, in order to verify the stability of the docked complexes over a reasonable time scale as well as to obtain more reliable binding free energy estimates, the docked complexes of the top hits and the reference inhibitor were subjected to 30 ns MD simulations and binding free energy calculations. The detailed binding mode of the top-ranked hits is presented using per-residue interaction energy decomposition as well as binding forces analyses.

To our knowledge, this report is the first account of an integrated computational approach applied in order to identify potential novel leads against HIV-Nef protein. The integrated *in-silico* computational and molecular bioinformatics tactics presented in this study could serve as a powerful tool for the identification of novel small molecules targeting HIV-Nef and many other emerging biological targets.

8.4. Computational Methods

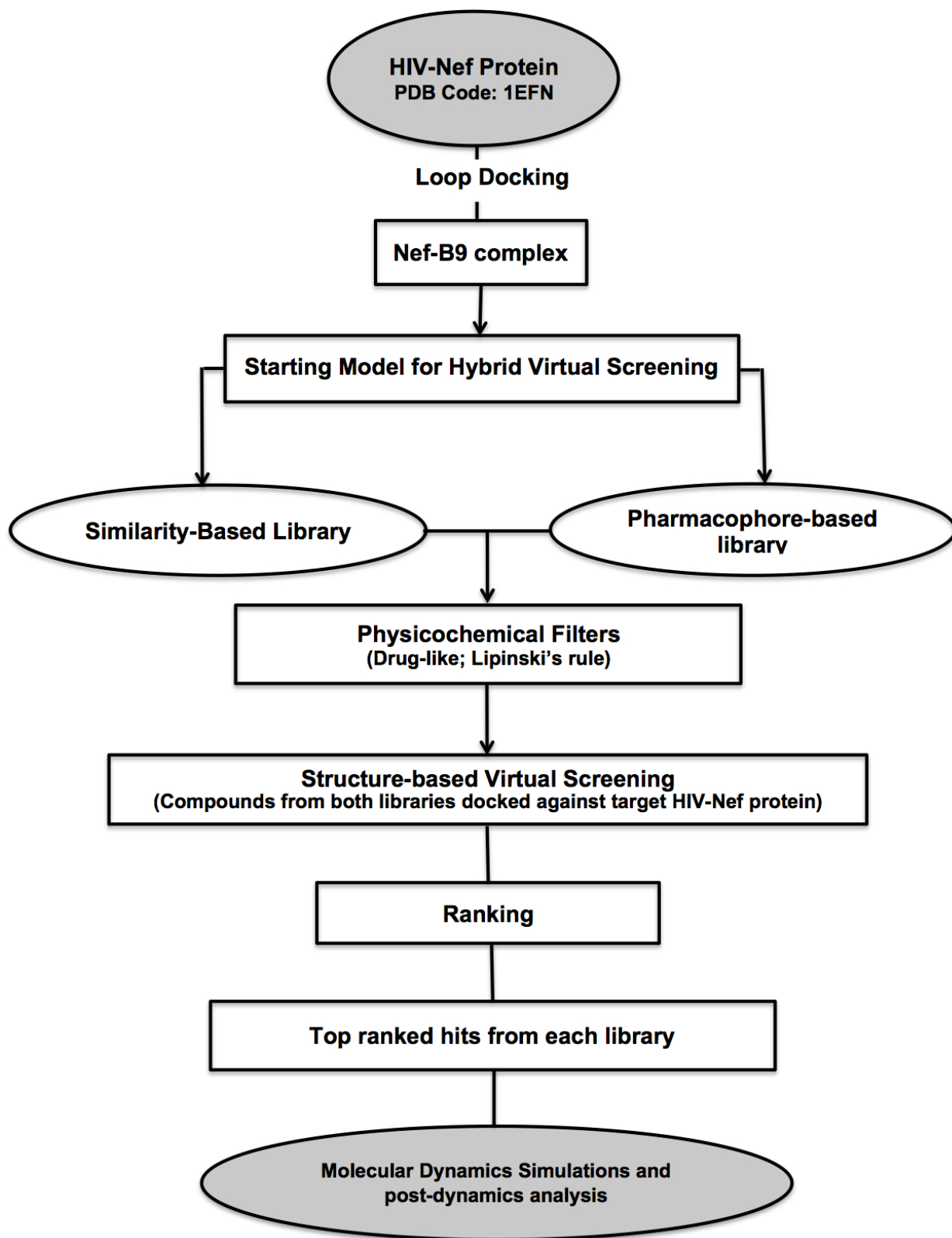The computational approached adapted in this study was outlined in **Figure 8.1**.

Figure 8.1. Flow diagram outlining the computational strategy adopted in this study.

8.4.1. B9 structure acquisition and preparation

The structure of newly discovered diphenylpyrazole moiety, B9 (**Figure 8.2.**), which proved as an effective antagonist of Nef-dependent HIV replication, was obtained from the DrugBank [33-35] [36] in the mol file format. Energy minimization and geometry optimization was carried out using MMFF94s force field integrated with Avogadro [36]. The optimized structure was saved in mol2 format for subsequent analysis.
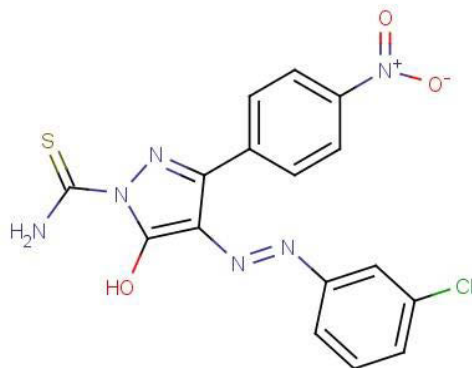

Figure 8.2. The 2D chemical structure of reference inhibitor as a prototype B9.

8.4.2. Protein systems

The crystal structure of conserved core of HIV-1 Nef dimer with SH3 domain (**Figure 8.3.**) was retrieved from protein data bank (PDB: **1EFN**). An entire comprehensive description of protein system preparation for docking and MD simulation is explained else in our previous reports [37-40].
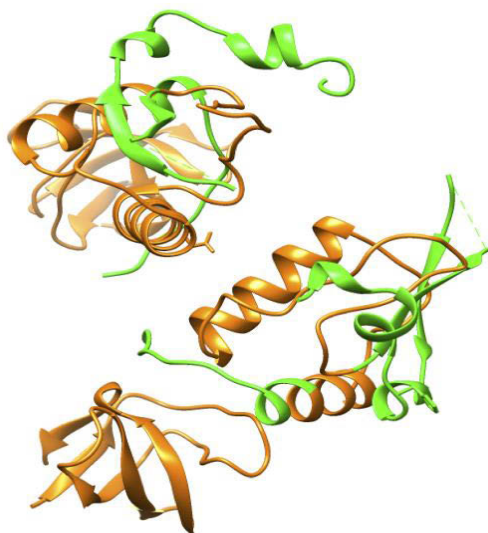

Figure 8.3. Crystal structure of HIV-Nef conserved region in complex with the SH3 domain (PDB: 1EFN). The residues at the SH3 domain ranging from 85-141 are highlighted in orange.

8.4.3. Ligand library Generation

8.4.3.1. Shape similarity based ligand library generation

The ZINC database [41], which contains over 21 million ready to dock compounds, was screened for the generation of the shape similarity-based compound library generated from the structural template of B9 (**Figure 8.2.**). The general chemical structure of B9 was drawn using Chemdraw (http://www.chemdraw.com/) and following this the structure was uploaded and the Zinc Database was queried for all scaffolds that having a >70% similarity. As explained in our previous reports [3], Lipinski rule of five [42] was implemented to increase drug-like properties amongst generated hits. The shape similarity search ultimately resulted in 560 drug-like hit molecules for further screening. Structure of all the query generated molecules were downloaded in a single mol2 file and were separated using Molegro Molecular Viewer (MMV) [43] and converted to the required pdbqt format using Raccoon for subsequent virtual screening using Autodock Vina [44].

8.4.3.2. Pharmacophore-based ligand library generation

The pharmacophore-based ligand library was generated using ZINC Pharmer [45]. The B9 structural template was uploaded into ZINC Pharmer in order to generate pharmacophoric regions, which contains fundamental points necessary for protein-ligand interactions as depicted in **Figure 8.4**. To produce a concise library with drug-like compounds and exclude duplication, Lipinski's Rule of Five[42] was set as a criterion. This extensive pharmacophore-based query led to the generation of 650 hits. Subsequently, structure of all generated were downloaded and processed as explained in the previous **section 8.4.3.1**.
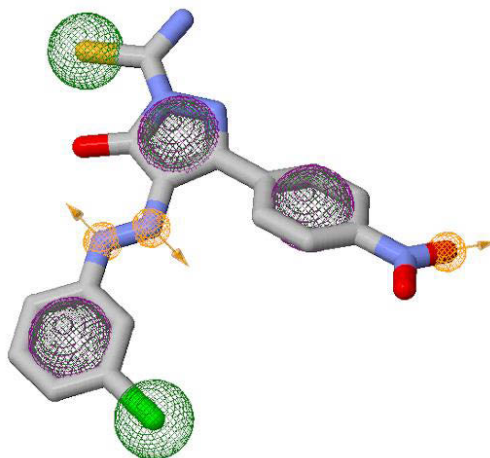
Figure 8.4. Pharmacophoric features used to search for new leads (yellow, green depicts hydrogen bond acceptor and hydrophobicity respectively, whereas pink depicts the aromatic moieties; arrows indicate the constraint direction).

8.4.4. Docking calculation: Hybrid virtual screening approach

In this study, the drug-like libraries generated using shape similarity and pharmacophore based approach were subjected to structure-based virtual screening using an enhanced loop docking protocol[46, 47] against the HIV-Nef target protein using Autodock Vina [48], using compound B9 as a proto type [49].

Virtual screening was carried out using Autodock Vina's default settings with an exhaustiveness set at 8.0. The grid box was defined around the following key residues Arg72, Gln151, Lys219, Met184, Ala114, Asp110, Asp113, Tyr115 and Phe116, which covered the conserved core region[14, 15, 50] of HIV-Nef protein. The X, Y and Z centre values of the grid box were defined as -106.28, -125.77 and -57.61, respectively, whereas, the X, Y and Z size dimensions were defined as 16, 13 and 20.4, respectively. All generated screening results are saved in the pdbqt format by Autodock Vina. Ten top-docked compounds on the basis of their binding affinities (kcal/mol) were selected from each of the two compound libraries. The different docked poses of the selected hits were visualised and selected using the ViewDock built-in plugin in Chimera.

8.4.5. Molecular Dynamics (MD) simulations and post-dynamic analysis

In recent year's molecular dynamics study have been extensively used to validate results generated from virtual screening as well as to gain insight into protein ligand interaction and stability of docked complexes. The top two best-docked compounds selected from the pharmacophore-based and shape similarity-based libraries were subjected to an all atom unrestrained MD simulations using GPU version of PMEMD engine provided with the Amber 12 software package [51]. The protein structure was represented and parameterized using the ff99sb variant of the AMBER force field implemented with Amber12 [52]. The geometry and the charges of the ligands were optimized at HF/6-31G* level using Gaussian 09 [53]. Finally ANTECHAMBER module was used to add partial atomic charges using restrained electrostatic potential procedure (RESP). The Leap module integrated with Amber12 was used to add missing hydrogen atoms and counter ions for neutralisation. Finally both systems were immersed within an orthorhombic box of TIP3P [54] water molecules such that no protein atom was within 8 Å of any box edge. Long-range electrostatic interactions were treated using particle mesh Ewald (PME) method with a direct space and vdW cut-off of 12 Å. Each system was initially minimized with an applied restraint potential of 500 kcal/mol $Å^2$ to the solute, followed by 1000 step of steepest descent followed by another 1000 steps of conjugate gradients. A further 1000 steps of minimization were carried out using conjugate gradients algorithm. All minimizations were carried out using CPU version of PMEMED engine provided with Amber12. Prior to minimization, a gradual heating from 0 to 300K was performed for both the systems with a harmonic restraint of 5kcal/mol $Å^2$ and a langevin thermostat with collision frequency of 1/ps using NVT ensemble. Both systems were subsequently equilibrated at 300K in a non-restrained NPT ensemble with pressure maintained at 1bar. The SHAKE algorithm [55] was used to constrain the bonds of all hydrogen atoms, a 2 fs time step and SPFP precision model was used for both MD runs. Finally, an isothermal isobaric (NPT) ensemble and Berendsen barostat [56] with a target pressure of 1 bar and a pressure coupling constant of 2 ps was used to perform 30 ns MD runs for both systems. The trajectories were saved in every 1ps and further analysis e.g. RMSD, RMSF, potential energy kcal/mol) were carried out using the PTRAJ module implemented with Amber12.

All system preparation, visualization and plots were carried out using Chimera molecular modelling tool and Origin data analysis software respectively.

8.6. Thermodynamic Calculations

The binding free energy profiles of the best-docked compounds from each compound library in complex with the HIV-Nef dimer were calculated using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) approach [41, 57-59]. We applied a single trajectory tactic and the binding free energy calculations were averaged over a timescale of 1000 snapshots at 30 ps intervals for the 5ns MD trajectory. The following set of equations describes the calculations of the binding free energy:

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

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

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

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

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

Where:

Symbols		Description
E_{gas}		Gas-phase energy
E_{int}		Internal energy
E_{ele}	and	Electrostatic and Van der Waals energies
E_{vdW}		Gaseous energy from the ff03 force field
E_{gas}		terms
G_{sol}		Solvation free energy
G_{GB}		Polar solvation contribution
G_{SA}		Nonpolar solvation contribution

In order to determine the individual amino acid contribution towards total binding free energy between the top two hits from virtual screening with HIV-Nef, a per-residue energy decomposition analysis of the interaction energy for each residue was computed by using the MM/GBSA binding free energy decomposition protocol implemented in Amber 12.

8.5. Results and Discussion

8.5.1. Validation of Computational Approach

Molecular docking study can often led to identification of false positives [60, 61]. To ensure the robustness of our docking protocol we compared the docking score (binding energy) of two previously reported Nef inhibitors with their respective biological activity. These datas were compared against the docking score and inhibitory activity of compound B9. To ensure the correct trend in binding free energies among the ligands used as a test set, MM/GBSA based binding free energy was calculated taking in account compound B9 as a standard. Recent findings suggest that the binding free energy profile from MM/GBSA protocol often highly correlates with the experimental activity[61-65]. **Table 8.1** highlights molecular docking and MM/GBSA based binding free energy profile of two Nef inhibitors (**Figure 8.5.**) used a test set in comparison with B9. Both of these compounds bind at dimeric active site of HIV-Nef (**Figure 8.6.**) with the binding free energies highly correlate with their respective biological activity. Outcome from this validation confirmed the robustness of our approach and its accuracy in identifying novel *in-silico* molecules targeting HIV-Nef.
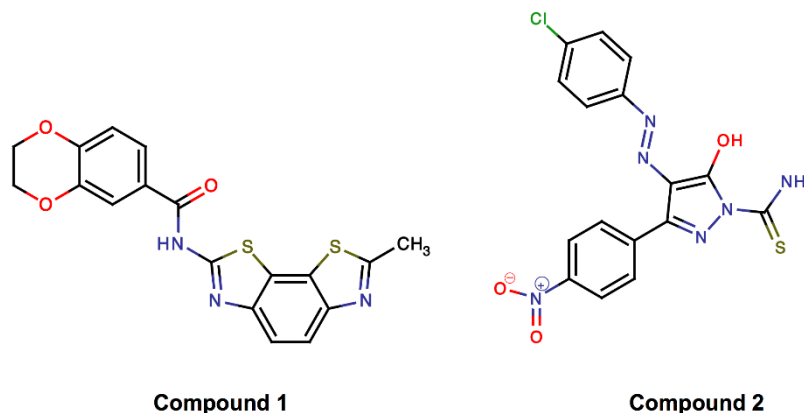

Compound 1   Compound 2

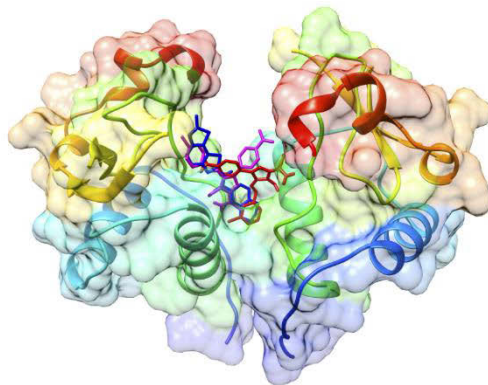**Figure 8.5. The 2-D structure of two HIV-Nef inhibitors used as a test set [14].**

Figure 8.6. The binding orientation of B9 (red), compound 1 (violet) and compound 2 (blue) inside the dimeric interafce of HIV-Nef.

Table 8.1. Comparison of biological activity, docking score and ΔG_{bind} among two test set compounds, compound 1 [14] and compound 2 [14] in comparison with B9.

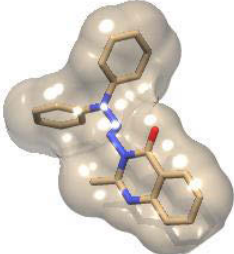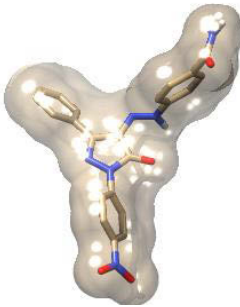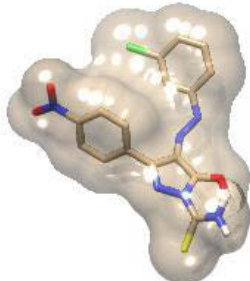| Inhibitors	IC$_{50}$ (μM)[14]	Docking Score (kcal/mol)	ΔG_{bind} (kcal/mol)*
Compound 1	4.8	-6.4	-16.0583±0.3032
Compound 2	5.4	-5.6	-12.3062±0.2042
B9	2.8	-6.8	-18.0694±0.4378

* Binding free energy profile using MM/GBSA approach
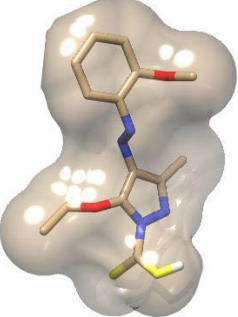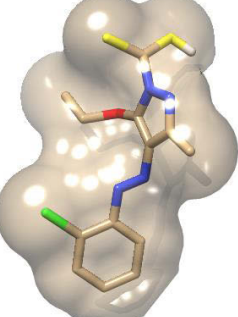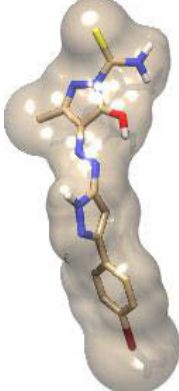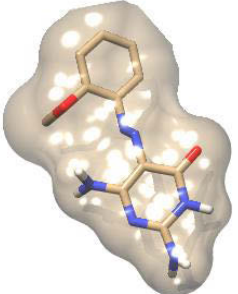
8.5.2. Hybrid Virtual screening

To date, only a single known molecule (compound B9) that inhibits the interaction of Nef exists. Therefore, the main objective of this study was to identify potential drug-like molecules targeting HIV-Nef, taking in account the structural and pharmacophoric features of compound B9 as reference template.
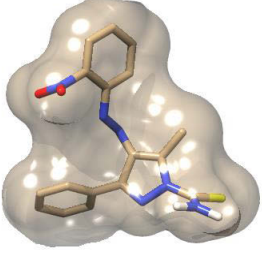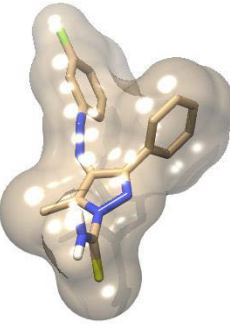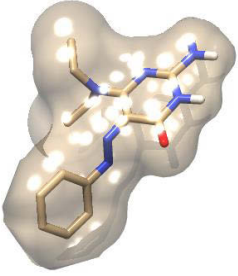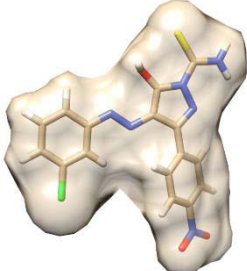
As a result of the extensive computational cost incurred, virtual screening has become quite feasible for the mining of large compound libraries. The ZINC database query for compounds sharing shape similarity to our diphenylpyrazole scaffold-containing template (B9) yielded 560 compound hits. Likewise, our query of the ZINCPharmer for pharmacophoric similarity yielded a total of 650 hits. In an effort to unveil potential new drug-like leads against the HIV-Nef target

protein, we merged both the generated structure-based and pharmacophore-based compound libraries obtained from virtual screening and elucidated the top 10 best-docked compounds based on binding energy (**Table 8.2.**).

Table 8.2. The top-ranked 10 hits leads ranked from highest to lowest according to binding affinities.

Library	Rank	ZINC ID	3D Structure	Binding energy (kcal/mol)	xlogP	H-bond Donors	H-bond Acceptors	Molecular weight (g/mol)
S*	1	ZINC04177596		-8.8	2.95	3	10	428.41
P**	2	ZINC36617540		-8.6	4.38	0	6	355.40
S	3	ZINC18055982		-8.4	4.51	3	9	402.82

P	**4**	ZINC17142282		-8.1	4.02	1	6	336.44
P	**5**	ZINC17142273		-8.0	4.64	1	5	340.86
S	**6**	ZINC18129984		-7.9	3.79	4	8	406.27
P	**7**	ZINC18009096		-7.8	1.22	5	8	260.23

S	8	ZINC17129910		-7.7	4.13	2	8	366.41
S	9	ZINC17129824		-7.6	4.36	2	5	339.40
P	10	ZINC25492082		-7.4	2.58	3	7	286.34
R		B9		-6.8	4.92	2	6	402.815

*top docked compound from shape similarity-based library, **top docked compound from Pharmacophore-based library, R= reference inhibitor

It was observed that all of the top 10 compounds possessed better binding affinities compared to the prototype, B9, with binding free energy ranging from -7.3 kcal/mol to -8.8 kcal/mol for shape similarity based library and -7.0 kcal/mol to -8.6 kcal/mol for pharmacophore based library. A surprising observation was that, compounds shared shape similarity had lower binding free energies (kcal/mol) on average compared to pharmacophore based library (**Table 8.3.**). Though

these differences were very subtle but it raises question on the importance of inclusion of specific pharmacophoric features that could be important for binding to target.

Table 8.3. The top 10 ranked ZINC compounds from shape based and pharmacophore based library

Rank	Shape similarity-based library	Binding affinity (kcal/mol)	Pharmacophore-based library	Binding affinity (kcal/mol)
1	ZINC04177596	-8.8	ZINC36617540	-8.6
2	ZINC18055982	-8.4	ZINC04565301	-8.4
3	ZINC18057425	-7.9	ZINC36617540	-8.1
4	ZINC18129984	-7.8	ZINC04557245	-8
5	ZINC17129824	-7.6	ZINC18009096	-7.8
6	ZINC17129910	-7.6	ZINC17140000	-7.3
7	ZINC17141898	-7.6	ZINC17140008	-7.2
8	ZINC17140002	-7.5	ZINC17140010	-7.1
9	ZINC17140006	-7.4	ZINC17140012	-7.0
10	ZINC04323473	-7.3	ZINC17140014	-7.0

Interestingly, the top 10 compounds from both libraries as well as compound B9 bind at the dimeric interface of HIV-Nef (**Figure 8.7.**).
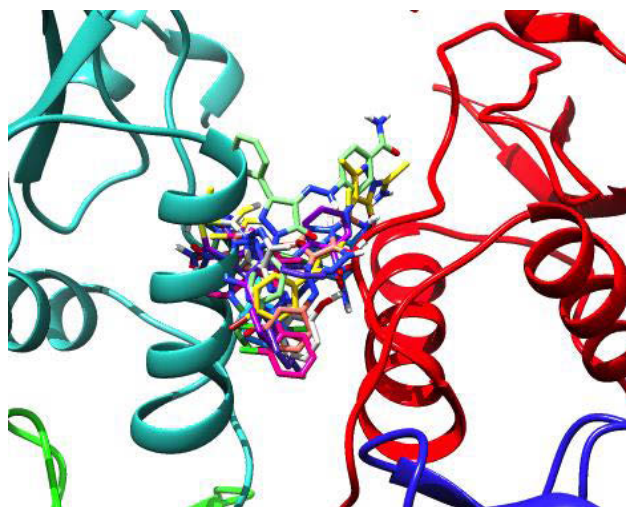

Figure 8.7. Outlined docking orientations of the top-ranked 10 compounds and prototype inhibitor, B9, docked into the active site.

It was perceived that all of the compounds, including B9, occupied the spaces between residues Arg72, Gln151, Lys219, Met184, Ala114, Asp110, Asp113, Trp113, Tyr115 and Phe116, respectively[14]. This was in great accordance to the fact that inhibition of Nef dimerization is crucial for inhibitory activity of potential inhibitors.

The top-docked compound from the pharmacophore-based library, **ZINC36617540**, was found to bind at the conserved core region of HIV-Nef dimer, interacting with the amino acid residues at SH3 domain (residues 85-141) (**Figure 8.13.**). Similar interactions with SH3 domain were observed in case of the top-docked compound from the shape similarity-based library, **ZINC04177596**, however, one hydrogen bond with Gln273 was observed (**Figure 8.13.**). Such interactions with the Nef-SH3 domain were anticipated to play a key role in inhibiting Nef dimerization, which will ultimately affect HIV-Nef functionality.
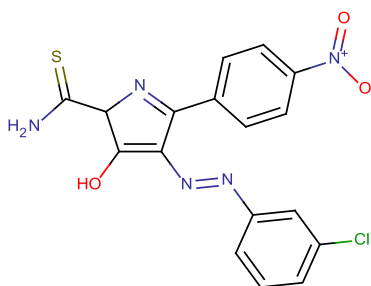
Compound B9
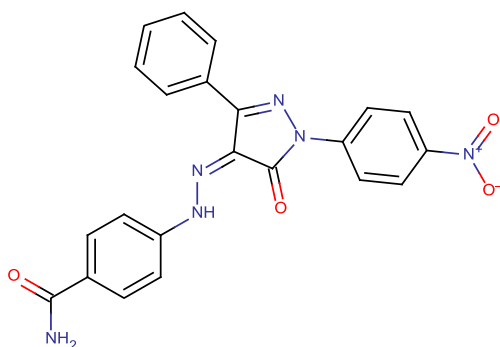
ZINC04177596
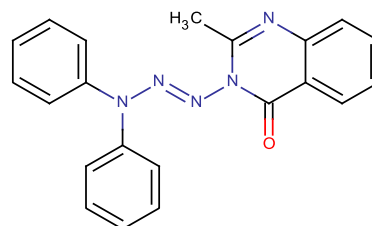
ZINC36617540

Figure 8.8. 2D structural comparison between reference structure of B9 with ZINC04177596 (top-docked compound from shape similarity-based library) and ZINC36617540 (top-docked compound from pharmacophore-based library), respectively.

8.5.3. MD simulations and Binding free energy calculations

Ligand orientation and/or binding energy results from docking calculations could be artifacts[32]. Therefore, it is necessary to supplement docking results with molecular dynamics simulations in order to validate the docking poses as well as binding energy values/trends. To this end, the two top-ranked screened compounds from each library, **ZINC36617540** and **ZINC04177596**, as well as diphenpyrazole containing reference drug (B9) complexed with the Nef protein were subjected to a 30 ns MD simulations. RMSD and potential energy fluctuations were monitored along the MD trajectories for all systems (**Figure 8.9.**). These findings provide further verification on the stability of our docked systems.
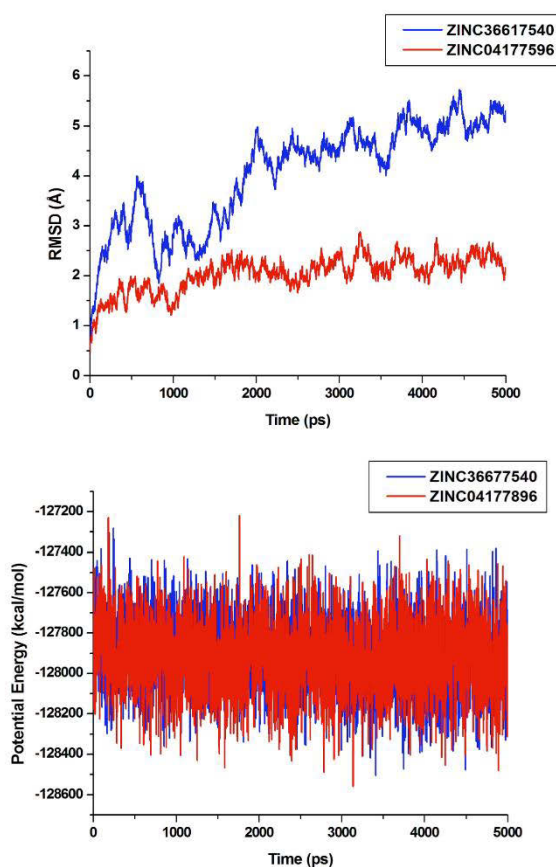

Figure 8.9. The RMSD and potential energy map of two top-ranked virtual screening hits (ZINC36617540 and ZINC04177596) complexed with Nef after 5ns MD simulation.

190

Figure 8.10 depicted the residue-based root mean square fluctuations (RMSF) of the two top-ranked system simulations. An interesting observation was that amino acid residues between 50-225 showed higher fluctuations in the pharmacophore-based (**ZINC36677840**) regions as compared to those residues within the shape similarity-based (**ZINC04177896**) regions. This might be due to the fact that these amino acids in the pharmacophore-based structure interacts less with the surrounding regions within the active site compared to the amino acids found in the shape similarity-based compound.
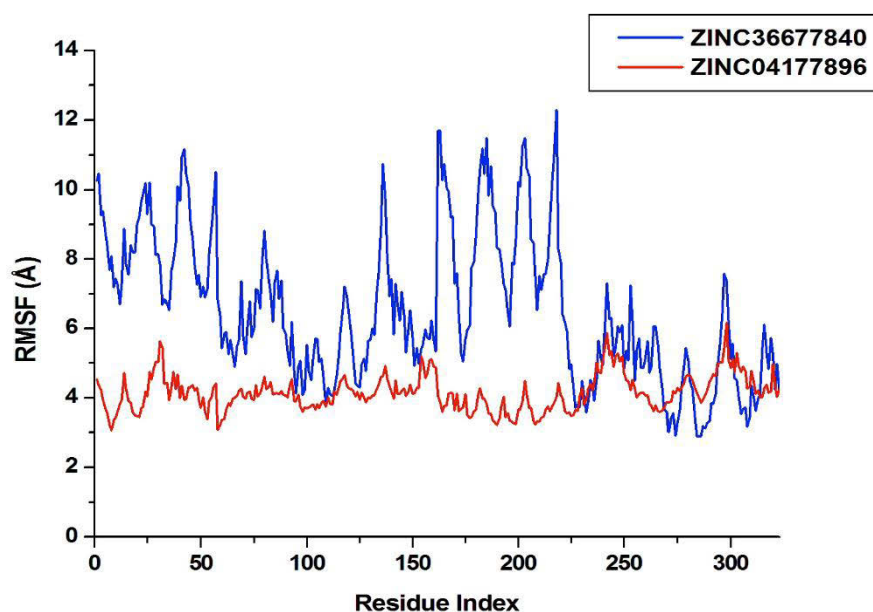
Figure 8.10. The per-residue fluctuations for the two top-ranked hits (ZINC36617540 and ZINC04177596) complexed with Nef dimer during the process of molecular dynamics.

All the free energy components are representative of averaged quantities over the 30 ns MD simulations calculated using the MM/GBSA approach are shown below in **Table 8.2**. Interestingly, the calculated binding free energy (ΔG_{bind}) of **B9** against Nef target protein is -18.06 kcal/mol compared to -20.28 kcal/mol and -28.75 kcal/mol in the case of **ZINC36617540** and **ZINC04177596**, respectively. As evident from the binding energy decomposition analysis (**Table 8.4.**), it was observed that ΔE_{vdW} and ΔE_{ele} are contributing the most towards ligand binding. It was also found that, the van der Waals contributions (ΔE_{vdW}) to the binding energy in

the pharmacophore-based (**ZINC36617540)** and shape similarity-based bound (**ZINC04177596)** Nef systems (-33.50 kcal/mol and -38.03 kcal/mol) are higher than that for the B9 bound Nef system (-26.90 kcal/mol). However, the calculated electrostatic contributions (ΔE_{ele}) to the binding free energy in case of pharmacophore-based (**ZINC36617540)** and shape similarity-based (**ZINC04177596)** bound with Nef systems were 0.18 kcal/mol and -2.40 kcal/mol lower respectively compared to -0.75 kcal/mol of the **B9** bound Nef system (**Table 8.4.)**.

Table 8.4. The binding free energies of the best-docked compounds from each compound library in complex with the Nef protein calculated using the MM/GBSA approach.

Ligand	ΔE_{vdw}	ΔE_{elect}	ΔG_{gas}	ΔG_{solv}	ΔG_{bind}
B9	-26.9049±0.5113	-0.7503 ±0.2415	-30.6537±0.6095	9.5843±0.2877	-18.0694±0.4378
ZINC36617540	-33.5008±0.3051	0.1831± 0.4217	-33.3177±0.5150	13.0906 ±0.4315	-20.2271± 0.2633
ZINC04177596	-38.0273±0.4274	-2.4075± 0.3093	-40.4348± 0.6440	11.6866±0.3009	-28.7482± 0.4112

8.5.4. Per-residue interaction energy decomposition analysis

To quantify individual amino acid interactions towards total binding free energy, we also computed per-residue interactions using MM/GBSA approach [66] (**Figure 8.11. and 8.1.2**) over 1000 snapshots from simulation. We noted from the per-residue interaction profile (**Figure 8.11. and 8.12.**) that Pro270, Asp259, Asp256, Gln273, Gln255 and Gln252 were found to be conserved interacting amino acid residues in with both the pharmacophore-based (**ZINC36617540)** and shape similarity-based bound (**ZINC04177596)** hits. However, major contributions seemed to be coming from Gln273, Leu99, Asn113, Gln252 and Asp256 in case of **ZINC04177596**, whereas in case of **ZINC36617540**, Pro109, Asp259, Pro270 and Gln273 were found to be contributing the most.
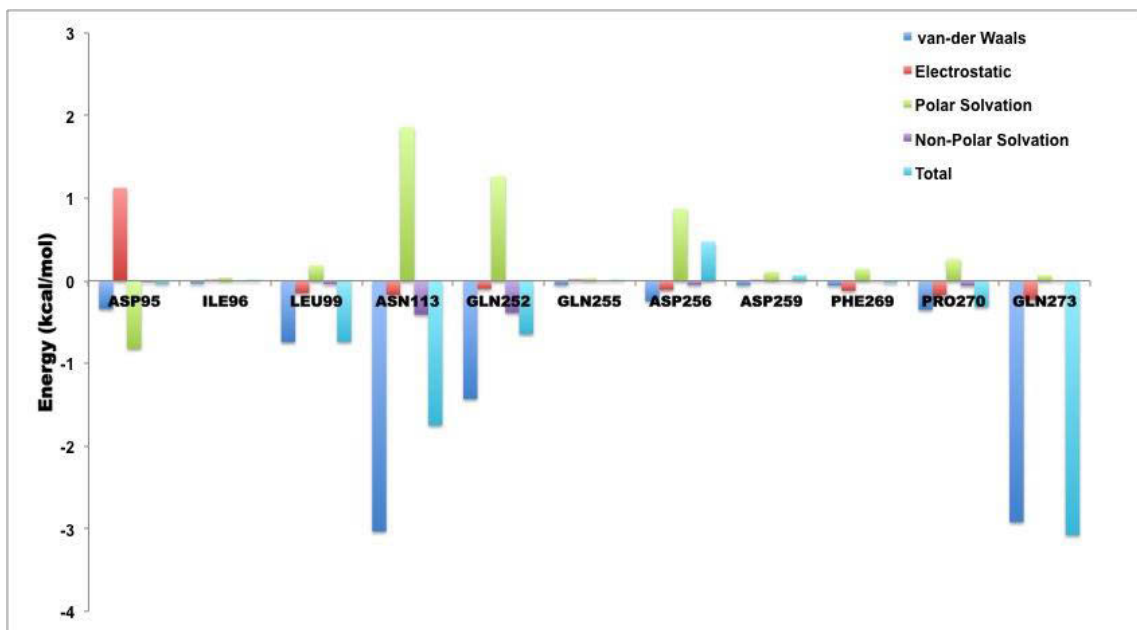
Figure 8.11. Per-residue energy decomposition for the top docked compound from the shape similarity-based library, ZINC04177596.
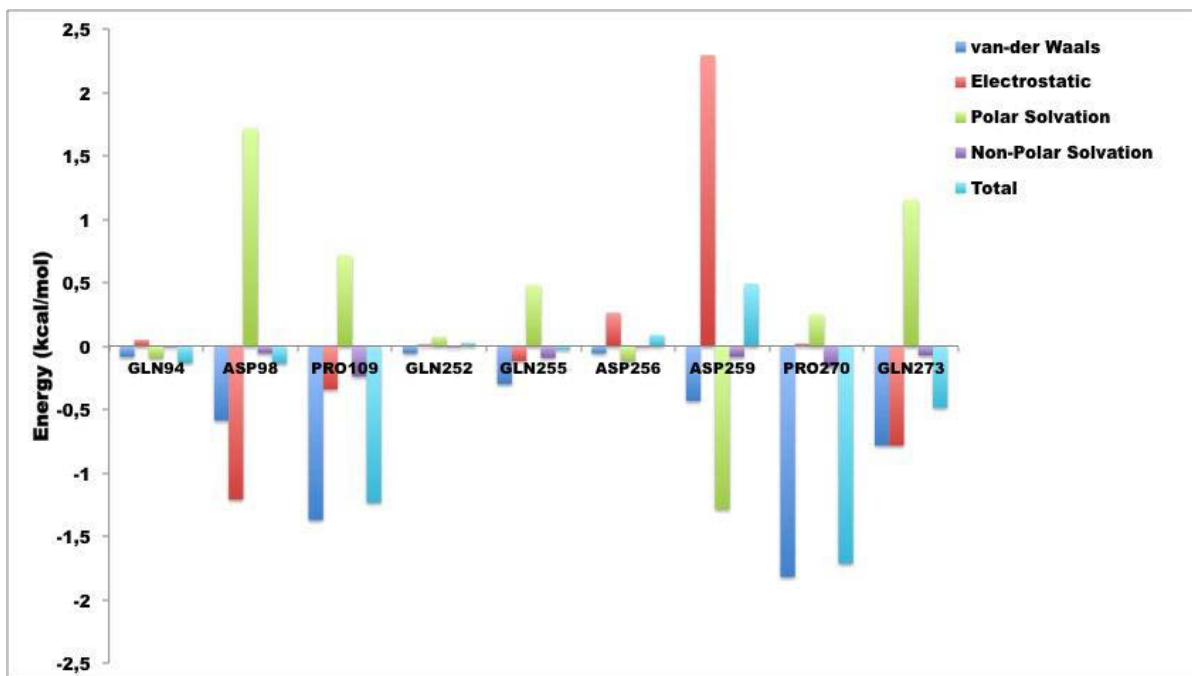

Figure 8.12. Per-residue energy decomposition for the top-docked compound from the pharmacophore based library, ZINC36617540.

It was also interesting to observe that Gln273 forms a hydrogen bond (average hydrogen bond distance 3.19 Å) with ZINC04177596, which could explain the high binding contribution from the Gln273 (**Figure 8.11.**). It is worth mentioning that all the above-mentioned residues that are involved in the interaction with the identified hits are located at HIV-Nef SH3 core domain. Thus, the identified hits are anticipated to play a key role in inhibiting Nef dimerization, which will ultimately affect HIV-Nef functionality.
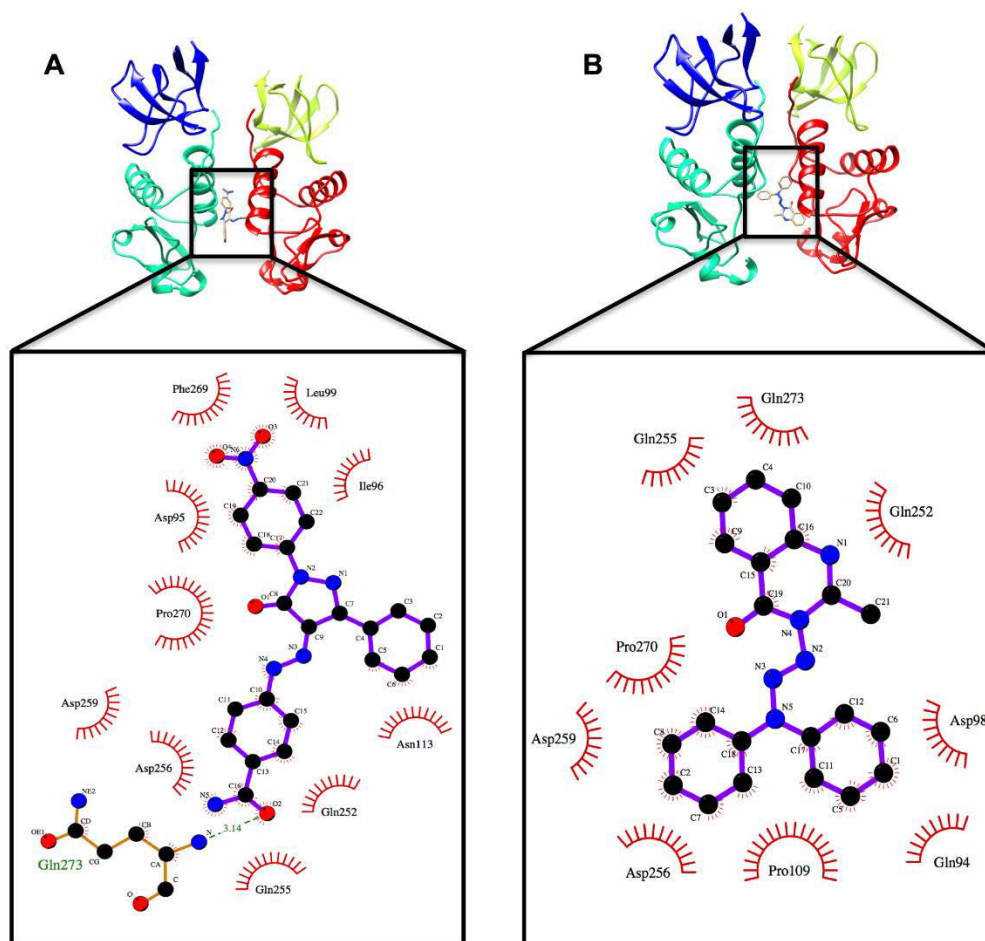
Figure 8.13. Depiction of the hydrogen bond and electrostatic interactions for the complexed shape similarity-based compound (ZINC04177596) with Nef (A), and pharmacophore-based compound (ZINC36617540) with Nef (B), respectively. The plots were generated using LigPlot package.

8.6. Conclusive structural features of potential Nef inhibitors

To add a further dimension to this study, we believed that informative chemical structural criteria for compounds to exhibit potential Nef dimerization inhibition should be presented herein. To

this end, we developed a combined pharmacophore map for the two top-ranked identified hits, **ZINC04177596** and **ZINC36617540**, (as explained in section **8.4.2**) to identify the crucial pharmacophoric features (**Figure 8.14.**). It is worth noting that the pharmacophoric groups were created based on the target-bound confirmations – target-bound confirmations proved to produce more reliable results as elaborated in our previous report[16]. Information from per-residue energy decomposition analysis and 2-D interactive ligand interaction plots (**Figure 8.15 B and C**) were also considered to identify these crucial residues surrounding the common pharmacophoric moieties. With this, we concluded a set of chemical structural features, as depicted in **Figure 8.15D**, which could be crucial for the activity of potential Nef inhibitors. Such information could serve as a guide or a route map in the design of potential Nef inhibitors as anti-HIV compounds.
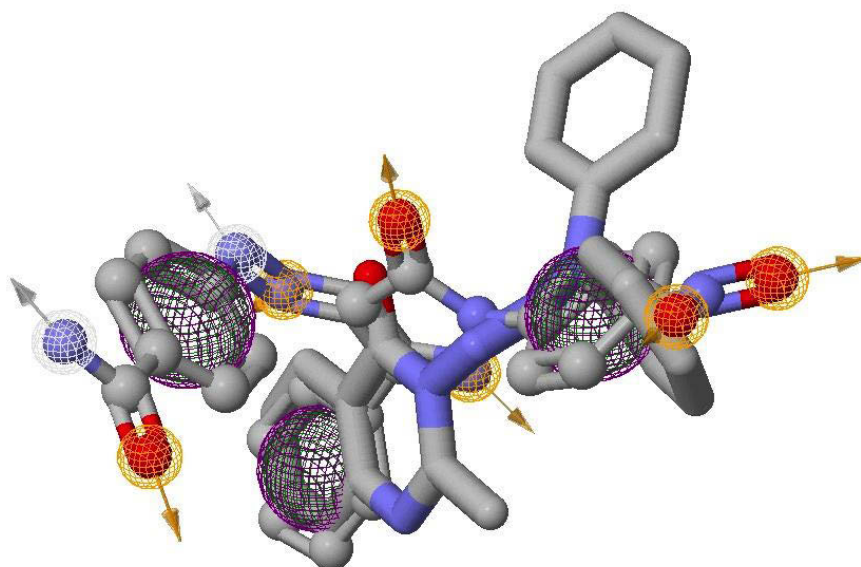
Figure 8.14. Common pharmacophoric features of top ranked compounds from both pharmacophore based and shape similarity-based library. The two hits, ZINC04177596, ZINC36617540 were aligned before creating the pharmacophoric maps. Yellow, green, pink, white depicts hydrogen bond acceptors, hydrophobic moiety, aromatic moiety, hydrogen bond donor respectively.
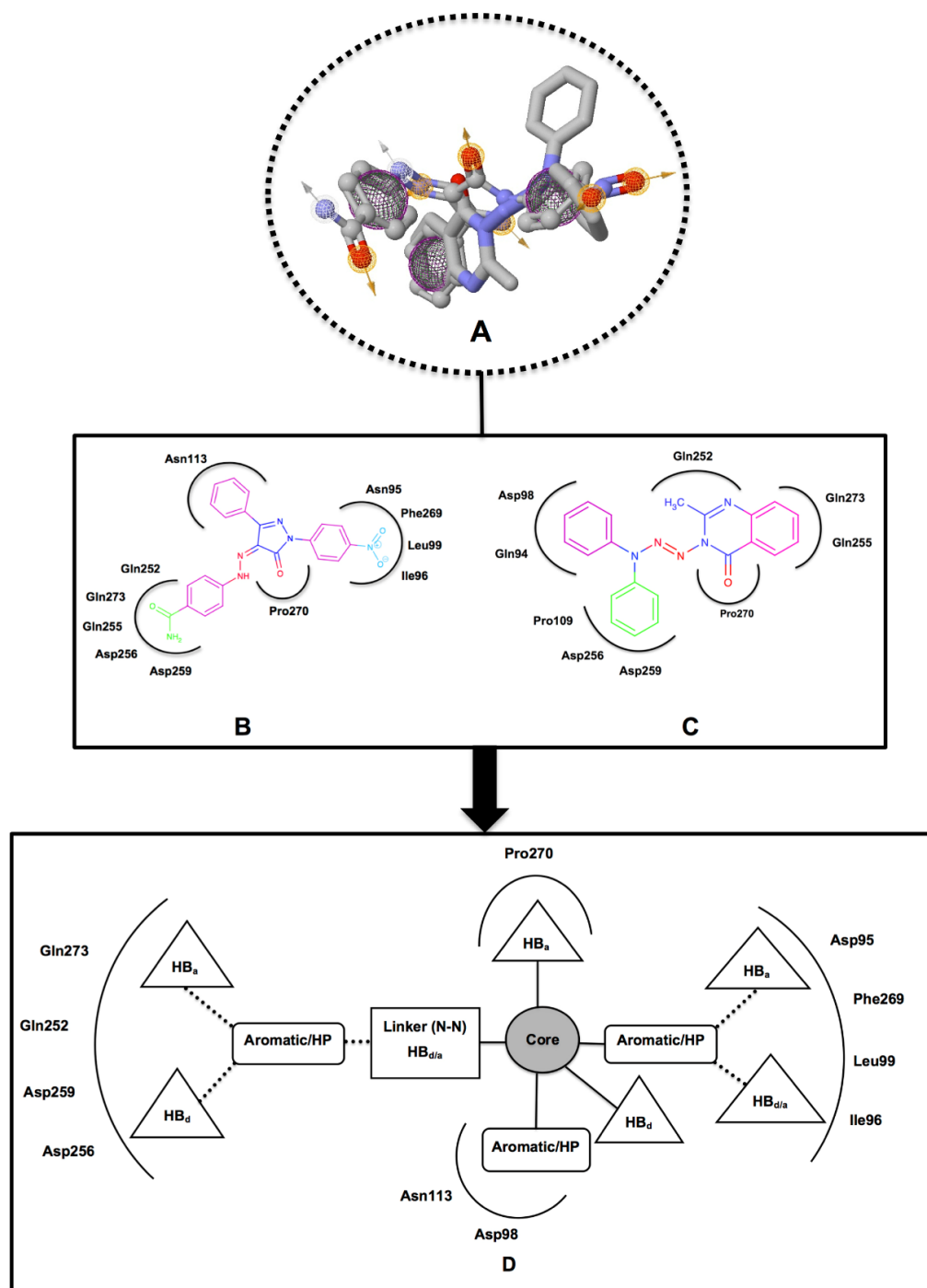
Figure 8.15. A Schematic representation showing the combined pharmacophoric features of ZINC04177596, ZINC36617540, A; 2-D interactive representation of ZINC04177596 and ZINC36617540 with Nef protein, B and C, respectively, and proposed chemical structural criteria of potential Nef inhibitors, C. HBa, HBd and HP denote for hydrogen bond donor, hydrogen bond acceptors and hydrophobic moieties, respectively.

8.7. Conclusion

In the present study, the structural features of compound B9, the only known HIV-Nef protein dimerization inhibitor, was used as a template to perform hybrid ligand-based virtual screening for the search for novel compounds with potential HIV-Nef inhibitory activity. The compounds library was generated based on two approaches; shape similarity- and pharmacophoric similarity-based virtual screening. Drug-like properties were used as filtering criteria for the search of the novel hits. This led to a result of 560 possible candidates from a shape similarity-based search and 650 possible candidates from the pharmacophore similarity-based search. The combined compounds library was then subjected to structure-based virtual screening using molecular docking to rank their binding affinities towards NEF protein. The best top-two candidates from each library were then selected for subsequent 5 ns MD and post-dynamic analysis. ZINC36617540 showed the best binding affinity from the pharmacophore-based library while ZINC04177596 showed the best from the chemical shape similarity-based library. Both compounds were found to bind prominently at the SH3 core domain of HIV-Nef. The calculated binding free energy (ΔG_{bind}) showed that the identified hits have better binding affinities by $\sim >$ 10 kcal/mol when compared to the prototype inhibitor, B9. Such significant difference in the binding affinities between the identified hits, ZINC36617540 and ZINC04177596 and B9 prompted us to propose a set of structural criteria that could be crucial for further optimization or design of novel Nef inhibitors. These sets of structural criteria were established based on a combined target-bound pharmacophoric maps of both hits as well as information gained from post-dynamic analysis.

However further experimental investigation would boost the current report, information highlighted herein could serve as a cornerstone strategy towards the design of novel potential inhibitors against HIV-Nef protein.

8.8. Acknowledgements

South Africa and the School of Health Sciences, UKZN for the scholarships and financial support.

8.9. Supplementary Material

The RMSD and binding free energy profiles of compound B9 during the simulation time were included as supplementary materials. The confirmation of HIV-Nef binding site and stability of binding site residues during simulation time were also included along with a detailed binding free energy profile of two test set molecules.

8.10. Conflicts of Interest

Authors declare no potential conflicts of interest.

8.11. References

1.   Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review in Biophysics and Biomolecular Structure 27*, 249-284.

2.   Xu, Y., Liu, H., Niu, C. Y., Luo, C., Luo, X. M., Shen, J. H., Chen, K. X., and Jiang, H. L. (2004) Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-yl)-butanes based on the structural modeling of human CCR5 receptor, *Bioorganic & Medicinal Chemistry 12*, 6193-6208.

3.   Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Drug Development Research 74*, 284-295.

4.   Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. . (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants, *Retrovirology* 1-23.

5. Morah, E. U. (2007) Are People Aware of Their HIV-positive Status Responsible for Driving the Epidemic in SubSaharan Africa? The Case of Malawi In *Development Policy Review*, pp 215-242.

6. Pani A, L., A. G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. . (2002) Targeting HIV: Old and New Players, *Current Drug Target -Infectious Disorders 2*, 17-32.

7. Kakkar, J., Chaudhary, K. K., and Prasad, C. V. S. S. (2013) Design of inhibitors using a combinatorial library for HIV-Nef and human SH3 domain interaction, *Bioinformation 9*, 777-781.

8. Fischl, M. A., Richman, D. D., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., Leedom, J. M., Groopman, J. E., Mildvan, D., Schooley, R. T., Jackson, G. G., Durack, D. T., and King, D. (1987) The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex - a double-blind, placebo-controlled trial, *The New England Journal of Medicine 317*, 185-191.

9. Trible, R. P., Narute, P., Emert-Sedlak, L. A., Alvarado, J. J., Atkins, K., Thomas, L., Kodama, T., Yanamala, N., Korotchenko, V., Day, B. W., Thomas, G., and Smithgall, T. E. (2013) Discovery of a diaminoquinoxaline benzenesulfonamide antagonist of HIV-1 Nef function using a yeast-based phenotypic screen, *Retrovirology 10*.

10. Landi, A., Iannucci, V., Van Nuffel, A., Meuwissen, P., and Verhasselt, B. (2011) One Protein to Rule them All: Modulation of Cell Surface Receptors and Molecules by HIV Nef, *Current Hiv Research 9*, 496-504.

11. Schaefer, M., Wonderlich, E., Roeth, J., Leonard, J., and Collins, K. HIV-1 Nef targets MHC-I and CD4 for degradation via a final common, *D - 101238921 4*, e1000131.

12. Vilaseca, J., Arnau, J. M., Bacardi, R., Mieras, C., Serrano, A., and Navarro, C. (1982) Kaposis sarcoma and toxoplasma-gondii brain-abscess in a Spanish homosexual, *Lancet 1*, 572-572.

13. Kakkar, J., Chaudhary, K. K., and Prasad, C. V. S. S. (2013) Design of inhibitors using a combinatorial library for HIV-Nef and human SH3 domain interaction, *Bioinformation 9*, 777-781.

14. Emert-Sedlak, L. A., Narute, P., Shu, S. T., Poe, J. A., Shi, H., Yanamala, N., Alvarado, J. J., Lazo, J. S., Yeh, J. I., Johnston, P. A., and Smithgall, T. E. (2013) Effector Kinase Coupling Enables High-Throughput Screens for Direct HIV-1 Nef Antagonists with Antiretroviral Activity, *Chemistry & Biology 20*, 82-91.

15. Smithgall, T. E., and Thomas, G. (2013) Small molecule inhibitors of the HIV-1 virulence factor, Nef, *Drug discovery today. Technologies 10*, e523-529.

16. Skelton, A. A., Maharaj, Y. R., and Soliman, M. E. S. (2014) Target-Bound Generated Pharmacophore Model to Improve the Pharmacophore-Based Virtual Screening: Identification of G-Protein Coupled Human CCR2 Receptors Inhibitors as Anti-Inflammatory Drugs, *Cellular and Molecular Bioengineering 7*, 45-57.

17. Unajak, S., Sawatdichaikul, O., Songtawee, N., Rattanabunyong, S., Tassnakajon, A., Areechon, N., Hirono, I., Kondo, H., Khunrae, P., Rattanarojpong, T., and Choowongkomon, K. (2014) Homology modeling and virtual screening for antagonists of protease from yellow head virus, *Journal of Molecular Modeling. 20*.

18. Chitrala, K. N., and Yeguvapalli, S. (2014) Ligand-based virtual screening to predict inhibitors against metastatic lymph node 64, *Journal of Receptors and Signal Transduction 34*, 92-96.

19. Saxena, S., Devi, P. B., Soni, V., Yogeeswari, P., and Sriram, D. (2014) Identification of novel inhibitors against Mycobacterium tuberculosis L-alanine dehydrogenase (MTB-AlaDH) through structure-based virtual screening, *Journal of Molecular Graphics & Modelling 47*, 37-43.

20. Srinivasan, P., Chella Perumal, P., and Sudha, A. (2014) Discovery of novel inhibitors for Nek6 protein through homology model assisted structure based virtual screening and molecular docking approaches, *The Scientific World Journal 2014*, 967873-967873.

21. Joshi, A. J., Gadhwal, M. K., and Joshi, U. J. (2014) A combined approach based on 3D pharmacophore and docking for identification of new aurora A kinase inhibitors, *Medicinal Chemistry Research 23*, 1414-1436.

22. Willett, P. (2006) Similarity-based virtual screening using 2D fingerprints, *Drug Discovery Today 11*, 1046-1053.

23. Polgar, T., and Keseru, G. M. (2011) Integration of Virtual and High Throughput Screening in Lead Discovery Settings, *Comb. Chem. High Throughput Screen 14*, 889-897.

24. Tanrikulu, Y., Krueger, B., and Proschak, E. (2013) The holistic integration of virtual screening in drug discovery, *Drug Discovery Today 18*, 358-364.

25. Hwang, S., Tamilarasu, N., Kibler, K., Cao, H., Ali, A., Ping, Y. H., Jeang, K. T., and Rana, T. M. (2003) Discovery of a small molecule Tat-trans-activation-responsive RNA antagonist that potently inhibits human immunodeficiency virus-1 replication, *The Journal of Bioogical Chemistry 278*, 39092-39103.

26. Zavodszky, M., Rohatgi, A., Van Voorst, J., Yan, H., and Kuhn, L. Scoring ligand similarity in structure-based virtual screening, *D - 9004580 22*, 280-292.

27. Feng, J., Guo, H., Wang, J., and Lu, T. (2014) Screening of Drug Target Proteins by 2D Ligand Matching Approach, *Chemical Biology & Drug Design 83*, 174-182.

28. Pirhadi, S., Shiri, F., and Ghasemi, J. B. (2013) Methods and Applications of Structure Based Pharmacophores in Drug Discovery, *Current Topics in Medicinal Chemsitry 13*, 1036-1047.

29. Cao, R., Li, W., Sun, H.-Z., Zhou, Y., and Huang, N. (2013) Computational chemistry in structure-based drug design, *Acta Pharmaceutica Sinica 48*, 1041-1052.

30. Lavecchia, A., and Di Giovanni, C. (2013) Virtual Screening Strategies in Drug Discovery: A Critical Review, *Current Medicinal Chemistry 20*, 2839-2860.

31. Drwal, M. N., and Griffith, R. (2013) Combination of ligand- and structure-based methods in virtual screening, *Drug discovery today. Technologies 10*, e395-401.

32. Shoichet, B. (2008) Hits, leads and artifacts from virtual and high-throughput screening, *Febs Journal 275*, 23-23.

33. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eisner, R., Guo, A. C., and Wishart, D. S. (2011) DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Research 39*, D1035-D1041.

34. Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Research 36*, D901-D906.

35. Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Research 34*, D668-D672.

36. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *Journal of Cheminformatics 4*.

37. Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E. S., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: A hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal of Medicinal Chemistry 46*, 3976-3985.

38. Makatini, M. M., Petzold, K., Sriharsha, S. N., Soliman, M. E. S., Honarparvar, B., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease, *Bioorganic and Medicinal Chemstry Letters 21*, 2274-2277.

39. Karpoormath, R., Sayed, Y., Govender, P., Govender, T., Kruger, H. G., Soliman, M. E. S., and Maguire, G. E. M. (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, *Bioorganic Chemistry 40*, 19-29.

40.   Makatini, M., Petzold, K., Alves, C., Arvidsson, P., Honarparvar, B., Govender, P., Govender, T., Kruger, H., Sayed, Y., JeronimoLameira, Maguire, G., and Soliman, M. (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane, *D - 101150203 28*, 78-88.

41.   Tsui, V., and Case, D. A. (2000) Theory and applications of the Generalized Born solvation model in macromolecular simulations, *Biopolymers 56*, 275-291.

42.   Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews 64*, 4-17.

43.   Hsu, M. C., Schutt, A. D., Holly, M., Slice, L. W., Sherman, M. I., Richman, D. D., Potash, M. J., and Volsky, D. J. (1991) Inhibition of HIV replication in acute and chronic infections invitro by a tat antagonist, *Science 254*, 1799-1802.

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

45.   Koes, D. R., and Camacho, C. J. (2012) ZINCPharmer: pharmacophore search of the ZINC database, *Nucleic Acids Research 40*, W409-W414.

46.   Blake, L., and Soliman, M. E. S. (2013) Bifunctional Anti-HIV/TB Inhibitors: Perspective from In-Silico Design and Molecular Dynamics Simulations, *Letters in Drug Design & Discovery 10*, 706-712.

47.   Moonsamy, S., and Soliman, M. E. S. (2014) Dual acting HIV inhibitors: integrated rational in silico design strategy, *Medicinal Chemistry Research 23*, 682-689.

48.   Trott, O., and Olson, A. J. (2010) Software News and Update 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.

49. Vivet-Boudou, V., Didierjean, J., Isel, C., and Marquet, R. (2006) Nucleoside and nucleotide inhibitors of HIV-1 replication, *Cellular and Molecular Life Sciences 63*, 163-186.

50. Narute, P. S., and Smithgall, T. E. (2012) Nef Alleles from All Major HIV-1 Clades Activate Src-Family Kinases and Enhance HIV-1 Replication in an Inhibitor-Sensitive Manner, *Plos One 7*.

51. 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.

52. 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.

53. 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*.

54. 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.

55. 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, *Journal of Computational Physics 23*, 327–341.

56. 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, *Journal of Chemical Physics 81*, 3684–3690.

57. 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.

58. 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.

59. 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.

60. Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations, *Drug Discovery Today 11*, 580-594.

61. Graves, A. P., Shivakumar, D. M., Boyce, S. E., Jacobson, M. P., Case, D. A., and Shoichet, B. K. (2008) Rescoring docking hit lists for model cavity sites: Predictions and experimental testing, *Journal of Molecular Biology 377*, 914-934.

62. Wichapong, K., Rohe, A., Platzer, C., Slynko, I., Erdmann, F., Schmidt, M., and Sippl, W. (2014) Application of Docking and QM/MM-GBSA Rescoring to Screen for Novel Myt1 Kinase Inhibitors, *Journal of Chemical Information and Modeling 54*, 881-893.

63. Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated From Docking, *Journal of Computational Chemistry 32*, 866-877.

64.   Rastelli, G., Del Rio, A., Degliesposti, G., and Sgobba, M. (2010) Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA, *Journal of Computational Chemistry 31*, 797-810.

65.   Lyne, P. D., Lamb, M. L., and Saeh, J. C. (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring, *Journal of Medicinal Chemistry 49*, 4805-4808.

66.   DeSantis, C., Siegel, R., Bandi, P., and Jemal, A. (2011) Breast cancer statistics, 2011, *CA: A Cancer Journal for Clinicians 61*, 409-418.

CHAPTER 9

Dynamic features of apo and bound HIV-Nef protein reveal the anti-HIV dimerization inhibition mechanism.
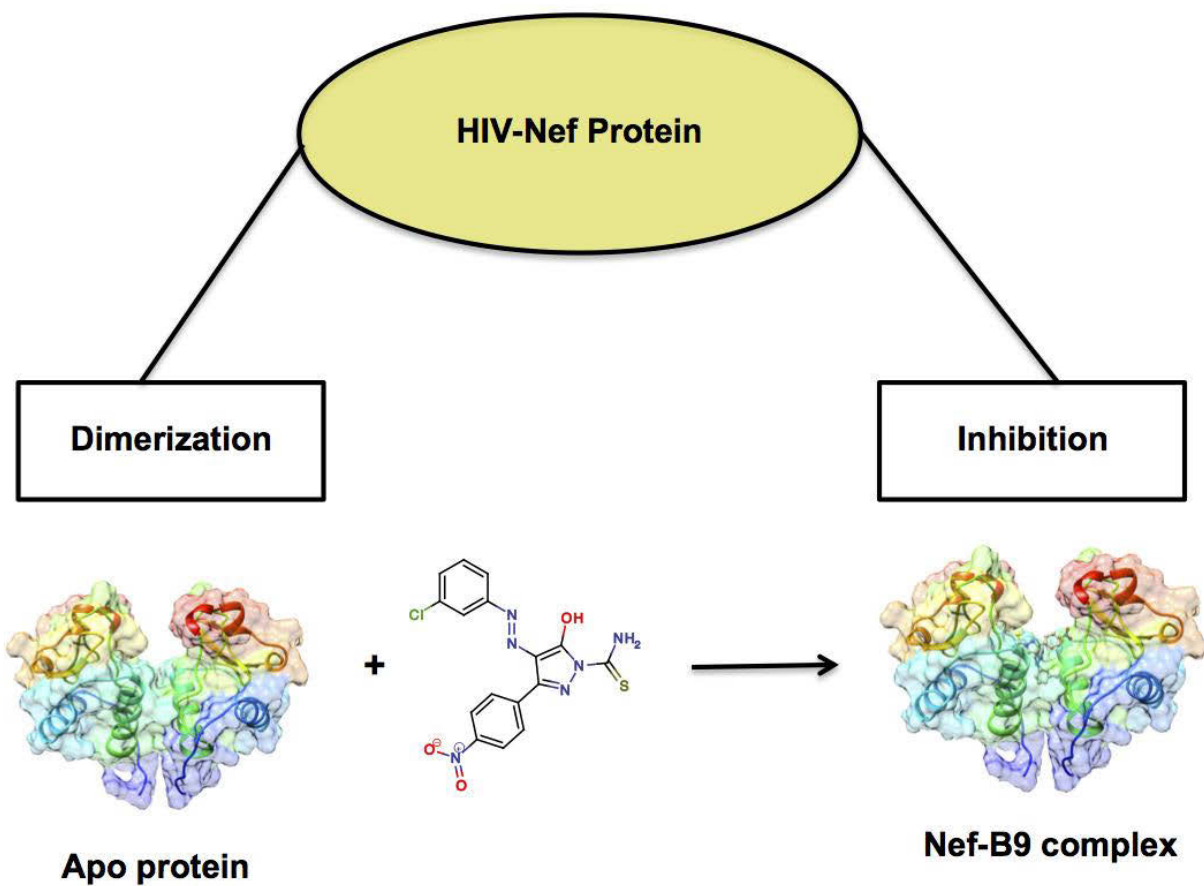
Suri Moonsamy[a], Soumendranath Bhakat[a], Mahmoud E. S. Soliman[a*]

[a] School of Health Sciences, University of KwaZulu-Natal, Westville Campus, 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

9.1. Graphical Abstract

9.2. Abstract

The first account on the dynamic features of Nef or negative factor, a small myristoylated protein located in the cytoplasm believes to increase HIV-1 viral titer level, is reported herein. The dimerization between two Nef subunits manipulates host cellular machinery and allows HIV to survive and replicate. Due to its major role in HIV-1 pathogenicity, Nef protein is considered an emerging target in anti-HIV drug design and discovery process. A Newly discovered compound B9 is believed to bind to the active site of HIV-Nef, which is located at the interface between two helical subunits, and interfere with protein dimerization. Here we report the first account on the dynamic features of HIV-Nef protein. In this study, comparative long-range all-atom molecular dynamics simulations were employed for apo and bound protein to unveil molecular mechanism of HIV-Nef dimerization and inhibition. Results clearly revealed that B9 binds at the dimeric interface of Nef protein and caused significant separation between residues involved in dimerization. Snapshots at a regular time interval along the time frame of molecular dynmaics in combination with increased inter-residue differences among orthogonally opposed residues, Asp108, Leu112, Gln104 confirmed a disruption in dimeric packing. Large differences in magnitudes were observed in case of radius of gyration (~1.5 Å), per-residue fluctuation (~2 Å), C-alpha deviations (~2 Å) which confirm a comparatively more flexible nature of apo conformation due to rapid dimeric association. Compared to the bound conformer, a more globally correlated motion in case of apo structure of HIV-Nef confirms the process of dimeric association. This clearly highlights the process of inhibition as a result of ligand binding. The difference in PCA scatter plot and per-residue mobility plot across first two normal modes further justifies the conformational flexibility and its inhibition in the presence of inhibitor.

The in-depth dynamic analyses of Nef protein presented in this report for the very first time would serve crucial in understanding its function and inhibition mechanisms. Information on inhibitor binding mode would also assist in designing of potential inhibitors against this important HIV target.

Keywords: HIV-Nef; dimerization; B9; molecular dynamics

Running title: HIV-Nef dimerization inhibition mechanism

9.3. Introduction

Since its emergence in early 1980s the "global killer" acquired immunodeficiency syndrome (AIDS) has already proclaimed the lives of estimated 22 million people worldwide [1, 2]. To date, the Human Immunodeficiency Virus type 1 (HIV-1), the instrumental agent in causing AIDS, is regarded as one the most challenging epidemic in the history of infectious diseases [2, 3]. According to different documented reports, an estimated 34 million individuals live with HIV/AIDS worldwide [4, 5]. Let alone in sub-Saharan Africa, an approximated 22.9 million individuals constitute the overall global estimate [4, 5]. Despite the continuous global effort, no cure yet exists for HIV/AIDS. Currently, the most effective therapeutic regimes consists of a multiple drug "cocktail", otherwise known as HAART, approved by food and drug administration (FDA), which consists of several antiretroviral, targeting different enzymes in HIV life cycle [6]. These drugs mainly include: protease inhibitors (PIs), integrase (IN) inhibitors reverse transcriptase (RT) inhibitors, entry inhibitors and progression inhibitors [7].

The development of resistant strains against almost all currently approved anti-retrovirals prompt researchers to find new drug targets, which can prove effective in reducing viral load from the host. The recently discovered HIV-Nef target, a small 27-35 kDa myristoylated protein, which plays a major role in HIV-1 pathogenicity [8]. Nef localizes primarily in the cytoplasm, but also partially to the plasma membrane and is one of many pathogen-expressed proteins, known as virulence factors. When undergoes dimerization in HIV, Nef protein is able to manipulate the host's cellular machinery and thus allow infection, survival or replication of the pathogen [9]. Nef stands for "Negative Factor" and even though the viral protein is a major component for HIV-1 replication in infected hosts, Nef markedly elevates viral titers [5]. As Nef is one of the accessory proteins of the HIV genome, thus it has a central role in the down regulation of host immunity.

Because of its major role in HIV-1 pathogenicity, Nef protein has been proved to be a very important target in anti-HIV drug design and discovery process. Understanding Nef dimerization process and its inhibition mechanism is crucial for the design of potent inhibitors as anti-HIV. To date, a handful of small molecule inhibitors were identified using high throughput screening (HTS) which block the Nef dependent HIV replication. Among them, a diphenylpyrazolodiazene containing small molecule inhibitor, B9 (**Figure 9.1.**) appeared to be specifically potent (IC50

2.8 µm) at blocking Nef activity by preventing dimerization of the two Nef subunits [10, 11]. The binding pocket of compound B9 with HIV-Nef dimeric interface proved as a future "hot spot" to target this protein[10].
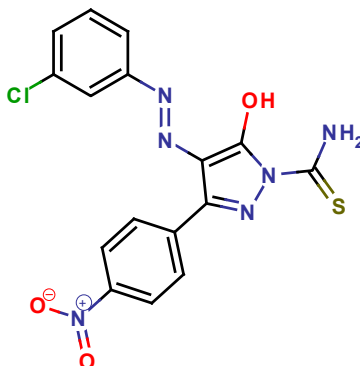

Figure 9.1. 2D structural representation of diphenylpyrazolodiazene containing Nef inhibitor, B9.

Molecular dynamics simulations and post dynamics calculations emerged as a close counterpart to experiment and helps in understanding the complex biological phenomena's. Application of long-range molecular dynamics simulations to unveil folding-unfolding behavior of biological macromolecules paved a way to understand the complex process of macromolecular dimerization [12-16]. Recent applications of enhanced post-dynamics approaches proved to be efficient in understanding the conformational landscape of biological macromolecules. Principal component analysis (PCA) or essential dynamics analysis is one of the widely used enhanced post-dynamics approaches to explore structural fluctuations among different biological systems [7, 17]. A large portion of overall fluctuation can often be accounted by a few low frequency eigenvectors with high Eigen values. If motion between two different macromolecules is similar then the Eigen vectors coming from individual trajectories should be similar to each other. For this reason, PCA has proved an efficient tool that can be used to compare motions of different macromolecules. Besides principal component analysis a number of post-dynamics analyses were also applied to understand the conformational behavior of biological systems [7, 18]. Dynamic cross correlation analysis is one of those techniques, which has been applied to understand the difference in macromolecular motion across different biological systems [19, 20].

In order to understand the dynamic landscape HIV-Nef dimerization process and its inhibition, comparative molecular dynamics simulations were employed for the apo and bound protein. A wide range of post-dynamic analyses were carried out in order to accomplish this task– these include; dynamic cross correlation (DCC), principal component analysis (PCA), radius of gyration (Rg), protein mobility plots as well as monitoring other several metrics.

To our best knowledge, this is the first account of such comprehensive computational study on this crucial HIV target. Therefore, we believe that this report serves as a cornerstone towards the understanding HIV-Nef protein structure and dynamics and its inhibitory mechanism. The compilation of computational and bioinformatics tools presented herein could also be implemented within the drug discovery and development of more potent HIV inhibitors against Nef.

9.4. Computational methods

9.4.1. Protein structure preparation

The crystal structure of HIV-Nef conserved core in complex with SH3 domain was retrieved from protein data bank (PDB: **1EFN**) for subsequent simulations [21]. The SH3 domain was deleted from the protein system to generate a native model of HIV-Nef core domain. The protein structure was prepared using Chimera[22] molecular modelling suite.

9.4.2. B9-Nef complex preparation

The recent diphenylpyrazole-based HIV-Nef inhibitor, B9, was believed to bind at dimeric region of HIV-Nef [10]. To generate an initial starting structure of B9-Nef complex, B9 was docked into the active site of HIV-Nef, which is located at the helical dimeric region [10, 11] of HIV-Nef (**Figure 9.2.**). Prior to docking, the ligand and protein systems were prepared as explained in our previous reports [23, 24]. Autodock Vina[25] was used to generate docked conformations of B9- HIV-Nef complex.
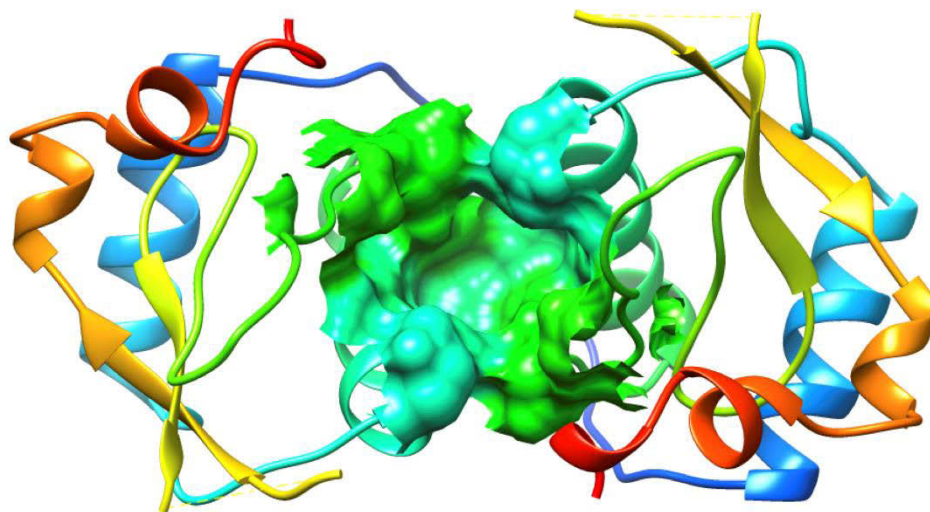
Figure 9.2. Graphical representation of HIV-Nef active site located at the junction of dimeric cleavage.

Table 9.1. The active site residues and GRID box dimensions used to dock compound, B9 inside HIV-Nef active site.

Protein System	Active site residues*	Grid Box dimensions
HIV-Nef conserved core domain	Gln104B, Gln107B, Asp108B, Asp111B, Leu112B, Pro122B, Asp123B, Trp124B, Gln125B, Asn126B, Gln104D, Gln107D, Asp108D, Asp111D, Pro122D, Asp123D, Trp124D, Gln125D, Asn126D	Number of Points X= 22.0 Å Y= 24.0 Å Z= 24.0 Å Centre Grid Box X= 44.0 Å Y= 18.0 Å Z= 37.0 Å

* B and D refers to dimeric chains of HIV-Nef conserved core domain

The active site of HIV-Nef conserved domain is highlighted in **Figure 9.2** and **Table 9.1** and was used to generate a grid box with a spacing set at 1 Å and an exhaustiveness of 8. The top docked conformation of compound B9 complexed with HIV-Nef was then visualized using ViewDock plugin[26] integrated with Chimera[22].

9.4.3. Molecular Dynamic (MD) Simulations

All-atom molecular dynamics simulations were performed on both apo and B9 bound conformations of HIV-Nef using Amber12 [7, 27]. The GPU version of PMEMD engine provided with Amber 12 was used to perform all molecular dynamics simulations. The RESP procedure was used to generate atomic partial charges and geometry was optimized using Gaussian 09 at the HF/6-31G* level[28]. The ANTECHAMBER module was used to generate atomic partial charges for the ligand using GAFF force field[29]. The ff99sb force field implemented with Amber 12 was used to describe the protein system[30]. The LEAP module integrated with Amber 12 was used to add missing hydrogens and heavy atoms and required counter ions to neutralize the system. Both the systems were immersed into an orthorhombic box with TIP3P[31] water molecules such that no atom was within 10 Å of any box edge. Periodic boundary conditions were enforced and long-range electrostatic interactions were treated using particle mesh Ewald (PME) method[32] with a direct space and vdW cut-off 12 Å. Prior to system preparations the minimizations, heating and equilibration steps were performed as described in our recent report[7]. Finally a 100 ns production run was performed in a NPT ensemble with target pressure of 1bar and a coupling constant of 2 ps.

The trajectory in both cases were saved and analyzed in every 1 ps. The PTRAJ and CPPTRAJ modules[33] integrated with Amber 12 were used for post-MD analysis e.g. RMSD, RMSF, Rg, distance, PCA. All visualizations and plots were carried out using Chimera[22]/VMD[34] and Origin data analysis tool respectively.

9.4.4. Dynamic Cross Correlation (DCC)

The dynamic cross correlation between the residue-based fluctuations during simulation was calculated using the CPPTRAJ module integrated with Amber 12. The following equation describes the DCCR as:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\left(\langle \Delta r_i^2 \rangle \cdot \langle \Delta r_j^2 \rangle \right)^{1/2}}$$

Where, i and j stands for ith or jth residue and Δr_i or Δr_j represents displacement vectors correspond to ith and jth residue. The cross correlation (C_{ij}) varies within a range of -1 to +1 with lower and upper-limit indicates a fully anti-correlated and correlated motion during simulation time. In this instance the DCC calculations were carried out taking in account backbone Cα atomic fluctuations.

9.4.5. Principle Component Analysis (PCA)

Principle component analysis (PCA) is widely used in recent years to reduce dimensionality of data obtained from molecular dynamics simulations to extract dominant modes responsible for conformational[7]. PCA was performed on C-alpha atoms using PTRAJ and CPPTRAJ modules[33] integrated with Amber 12. The data were averaged over 1000 snapshots with an equal interval of 100 ps in both cases. The first two principal components correspond to first two Eigen vectors of the covariance matrix[7]. The scatter plot showing the dominant conformational motion representative of each structure was created using Origin data analyses programme. The porcupine plots correspond to first two normal modes were created using ProDy interface[35] integrated with Normal Mode Wizard plugin of VMD[34].

9.5. Results and Discussion

9.5.1. Binding mode of B9 with HIV-Nef

The docked structure of inhibitor B9 inside the active site of HIV-Nef dimeric cleavage provided crucial information on ligand binding landscape inside the active site of HIV-Nef (**Figure 9.3.**). It can be observed that –NO2 group attached with the phenyl ring involved in formation of two hydrogen bond interactions with Gln104 and Gln107 residues of one subunit of HIV-Nef dimer.

Interestingly, compound B9 binds at the dimeric cleavage of HIV-Nef dimer possessing multiple common residues from both sub-units in its active site. The 2D interaction plot (**Figure 9.4.**) highlighted the position of inhibitor B9 in the HIV-Nef dimeric cleavage and common active site residues from each dimeric subunit. The binding mode clearly shows the presence of a hydrophobic/aromatic moiety, which is involved in proper binding of inhibitor in the dimeric groove, consists of some major active site residues involved in dimer packing e.g. Asp108, Pro122, Leu112. **Figure 9.4** further highlights the position of pharmacophoric features, which are necessary for a proper binding of inhibitor within residues involved in dimer packing e.g. Gln104, Asn126, Gln107. The detailed pharmacophoric feature described in one of our recent communication in combination with a recently reported structure activity relationship study[36] may provide structural benchmark to develop novel small molecule inhibitors using a rational approach[37].

It is believed that binding of compound B9 in the HIV-Nef dimeric cleavage leads to inhibition of HIV-Nef dimerization. The docked structure of compound B9 was further used to obtain a dynamic insight into the mechanism of dimerization inhibition by aid of long-range molecular dynamics simulation.
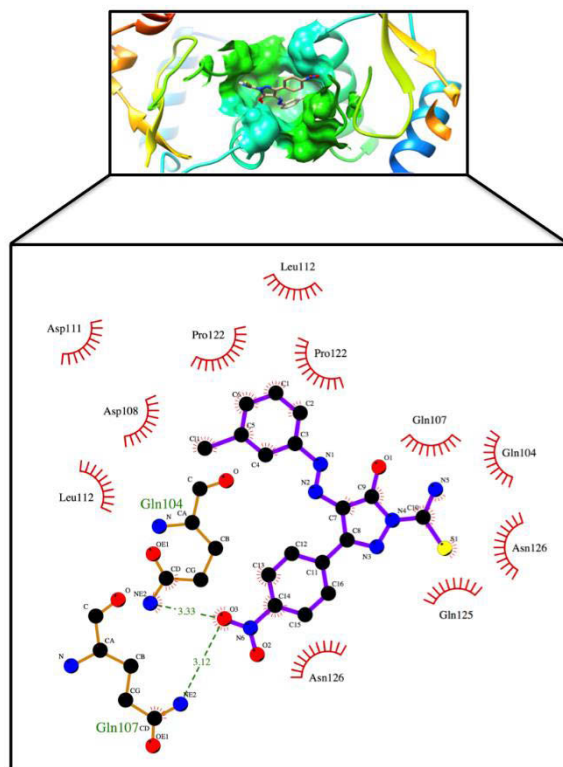
Figure 9.3. Residue interaction plot of compound, B9 inside the active site of HIV-Nef. Green dotted lines denote hydrogen bond interactions.
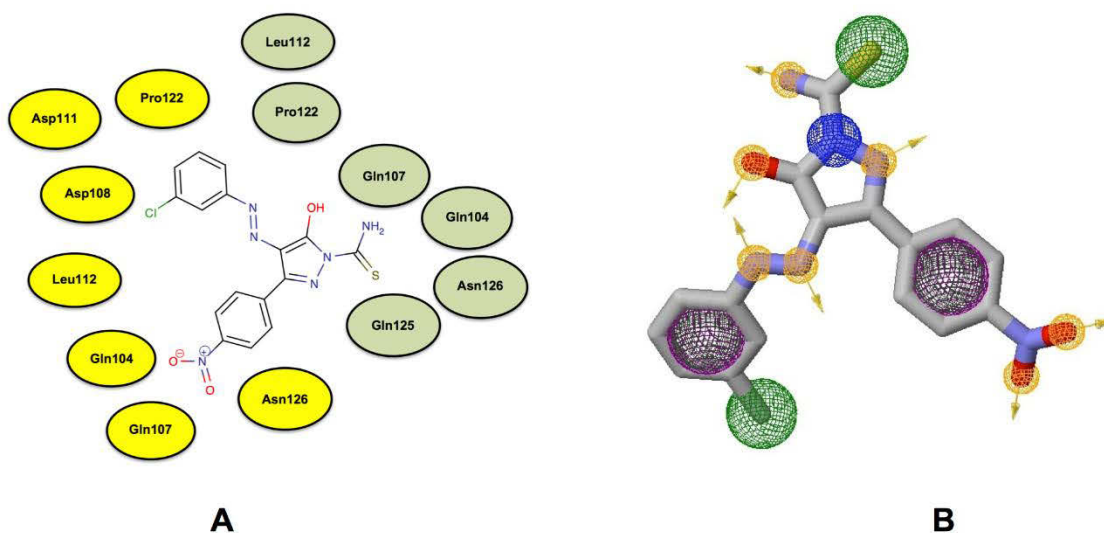

A

B

Figure 9.4. A. 2D interaction map of compound B9 at the dimeric cleavage of HIV-Nef. Yellow and light green indicates the location of residues in each helical subunit. B. Pharmacophoric feature of target bound conformation of compound B9.

9.5.2. MD simulations and post-dynamics analysis

9.5.2.1. RMSD, RMSF and radius of gyration (Rg)

Figure 9.5 highlights the time dependent root mean square deviation of backbone C-alpha atoms for both HIV-Nef free and ligand (B9) bound conformation. It was noticed that the backbone of both systems were well stabilized after a 30 ns time period. However, a larger magnitude (~2 Å) of fluctuation was observed in case of unbound conformation of HIV-Nef when compared with bound state. One possible explanation of this phenomenon is that the attachment of inhibitor molecules inside HIV-Nef active site leads to a conformational rigidity, which hampers conformational evolution of HIV-Nef, especially during the process of dimerization.
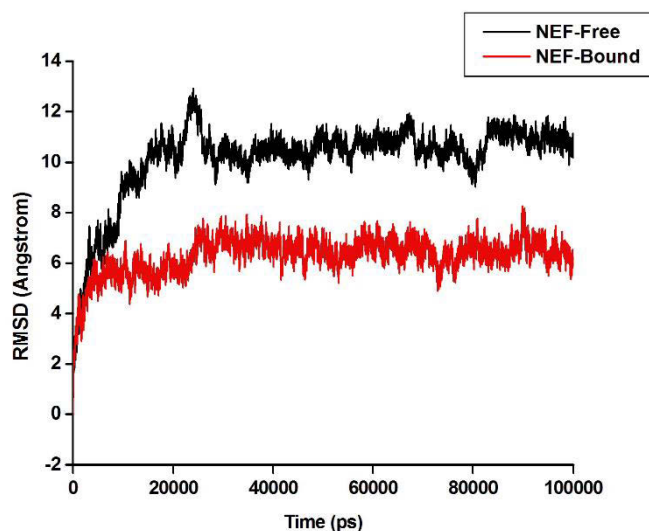

Figure 9.5. C-alpha backbone RMSD for HIV-Nef free and ligand bound conformations. The average C-alpha RMSD found to be 5.18 Å and 3.72 Å respectively for apo and B9 bound complex of HIV-Nef.

To assess the effect of inhibitor (B9) binding on the conformational flexibility of the HIV-Nef, per-residue RMSF of C-alpha carbons were computed for both ligand bound and apo conformations. From per-residue fluctuation (**Figure 9.6.**) it can be clearly noticed that presence of inhibitor inside the active site of HIV-Nef highly affected the conformational flexibility of the overall protein structure, which reflects the inhibition of dimerization in presence of inhibitor. The significant difference (~2 Å) in average per-residue fluctuation clearly indicates a conformationally flexible nature of apo conformation, which might account for the disruption in

dimerization. Furthermore, a more flexible region was observed between residues 138-152 which are located in one of the dimeric helix section of Nef (apo conformation). The higher value in RMSF might be due to the process of dimerization, which involves flexible helical residues located at Nef dimers. Whereas, the presence of B9 in the helical region inhibits the process of dimerization that affects the overall RMSF of Nef-B9 complex, as highlighted by lower average RMSF as well as lower magnitude of per-residue fluctuation at the helical region.
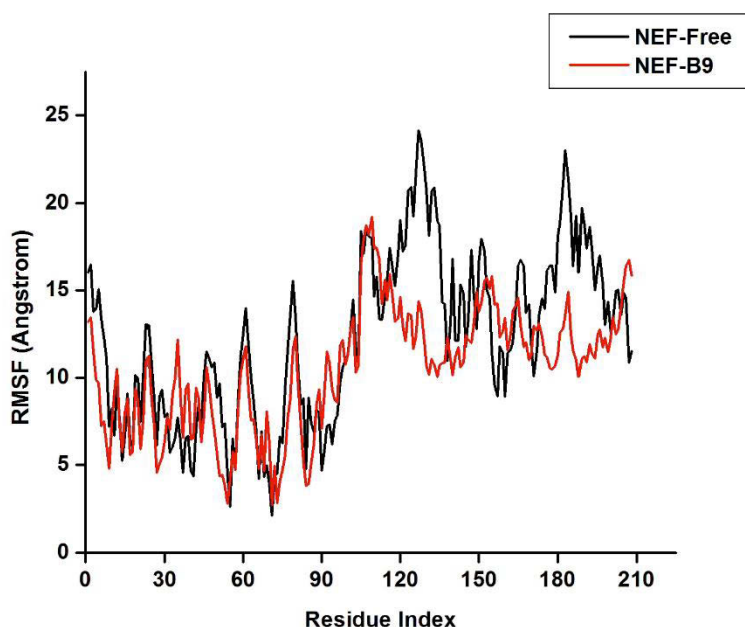
Figure 9.6. Residue based fluctuation of HIV-Nef free and inhibitor bound conformations of HIV-Nef during the simulation time. The average C-alpha per-residue fluctuation for apo and bound conformations were found to be 12.26 Å and 10.48 Å respectively.

The overall fluctuations in RMSF further correlate with C-alpha RMSD fluctuations, which confirm a larger overall fluctuation in case of free conformation when compared with ligand bound conformation of HIV-Nef. **Figure 9.7** highlights the deviation of C-alpha residues located at the helical region of HIV-Nef active site, which clearly indicates more flexible nature of residues in unliganded (apo) HIV-Nef system when compared with bound conformation. The apo conformation showed a higher value of C-alpha deviations at the Nef helical region with an average RMSD of 5.17 Å, whereas its B9 counterpart displayed a comparatively lesser deviation with an average RMSD of 3.71 Å. This significant difference in C-alpha RMSD's (~2 Å) between helical region of native apo and bound conformations further highlights the helical

flexibility, which played a significant role in the process of dimerization. Thus it can be further postulated that attachment of inhibitor B9 in the helical active site of HIV-Nef leads to conformational rigidity, consequently impairs dimerization process.
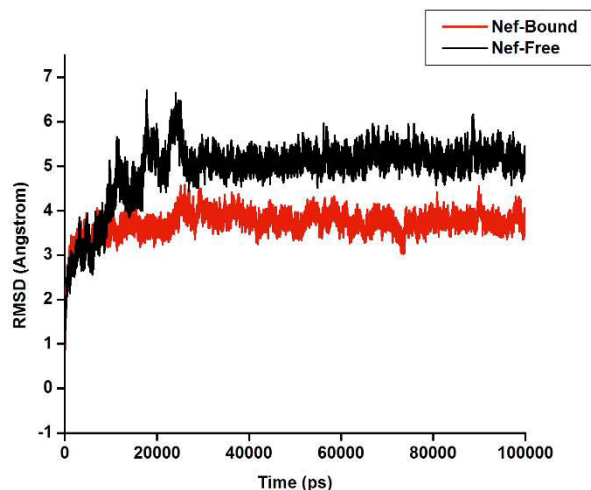

Figure 9.7. Deviation of C-alpha atoms of residues located at the active site helical region involved in the process of dimerization. The average RMSD's between apo and B9-Nef complex found to be 5.17 Å and 3.71 Å respectively.

These findings were further supported by the findings observed from the radius of gyration (Rg) (**Figure 9.8.**).
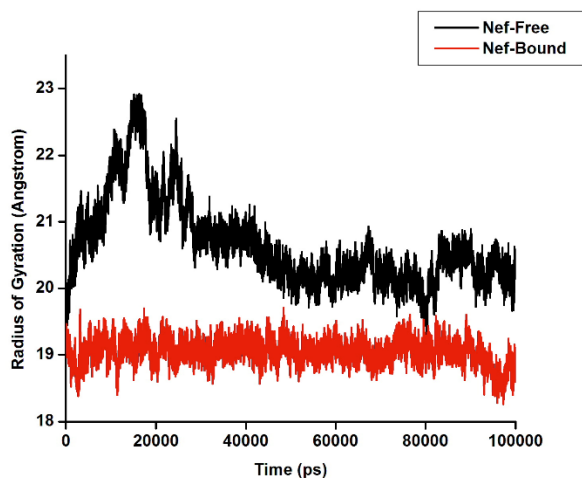

Figure 9.8. Radius of Gyration (Rg) of C-alpha atoms of HIV-Nef free and ligand bound conformation. The average Rg of apo and bound conformations found to be 20.64 Å and 19.04 Å respectively.

The shape and folding of Nef-bound and free conformations over the trajectory can be seen in terms of Rg. Throughout the simulation, the apo conformation of HIV-Nef showed comparatively higher Rg value as compared to B9-Nef complex. The native apo conformation displayed an average Rg of 20.64 Å whereas B9-Nef complex displayed a significantly lower Rg fluctuation with an average value of 19.04 Å. This larger breathing in Rg (~1.5 Å) in case of native apo conformation highly correlates with per-residue fluctuations and RMSD outcomes which justified an increased biomolecular flexibility of apo structure as compared to bound one. Binding of inhibitor B9 in the dimeric cleavage of HIV-Nef and its negative effect on dimeric flexibility confirms its inhibitory mechanism by hampering the overall biomolecular flexibility and dimer packing.

9.5.2.2. Understanding HIV-Nef dimerization

To further understand the dynamics of dimerization process, the distance between the orthogonally opposed residues[11] in each dimeric subunit are monitored along the dynamic simulations. These residues are highlighted in **Figure 9.9**.
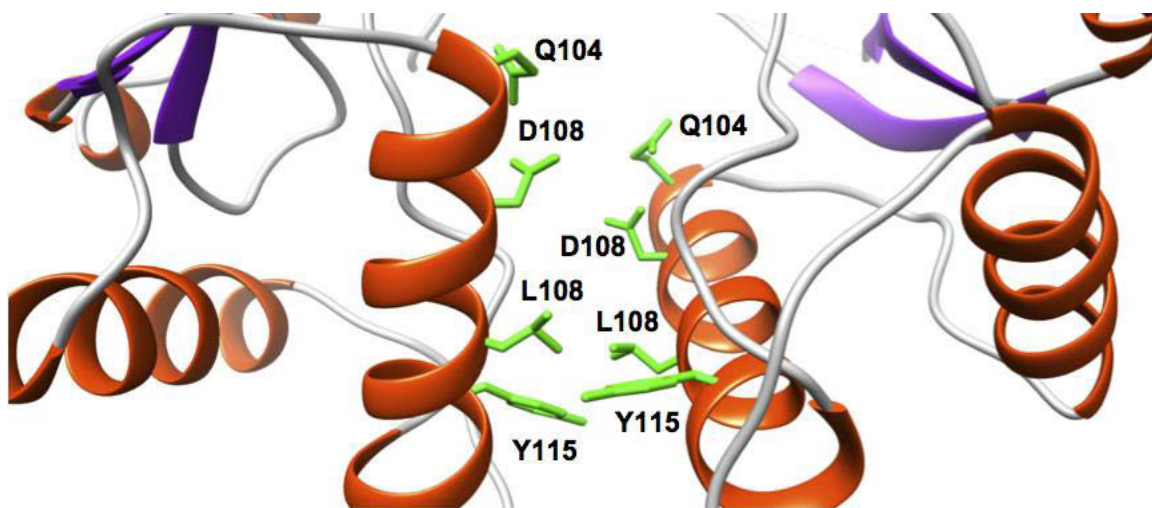

Figure 9.9. Position of orthogonally opposed residues at the dimeric helix of HIV-Nef believed to be involved in the process of dimerization[10, 11].

Figure 9.10, 9.11, 9.12 and **9.13** clearly suggests that in case of apo protein the distances between the orthogonal residues are significantly less when compared with bound conformation of HIV-Nef. Snapshots along the pathway of molecular dynamics simulations for apo and B9

bound conformations of HIV-Nef further justified a lack of dimerization in case of Nef-B9 complex (**Figure 9.14.**). The average distances between C-alpha atoms of orthogonally opposed Leu112, Gln104 and Asp108 residues in case of B9 bound conformation of HIV-Nef found to be 11.24, 24.23, 17.32 and 32.38 Å respectively. Whereas, in case of apo conformation the magnitude of average distances were comparatively less as the average C-alpha distances between Leu112, Gln104 and Asp108 found to be 10.93, 11.62 and 13.73 Å respectively. These significant differences in distances among residues involved in dimer packing confirmed lack of dimerization in Nef-B9 complex. This fact further co-relates with the experimental outcome, which confirmed a disruption of Nef dimerization by B9 using a cell based assay[10].

Molecular dynamics simulations further revealed that another oppositely placed helical residue present at the Nef dimeric region, Tyr115 moves further apart from each other in the B9 bound confirmation and thus confirming the inhibition in dimer packing as a result of disruption between two Nef helices (**Figure 9.12.**). The occupancy of compound B9 inside the active site of HIV-Nef dimer was found responsible in increasing the average distances among residues (**Table 9.2.**) involved in dimer packing, suggesting inhibition of HIV-Nef dimerization, which further substantiates the lack of conformational flexibility as a result of inhibitor binding.
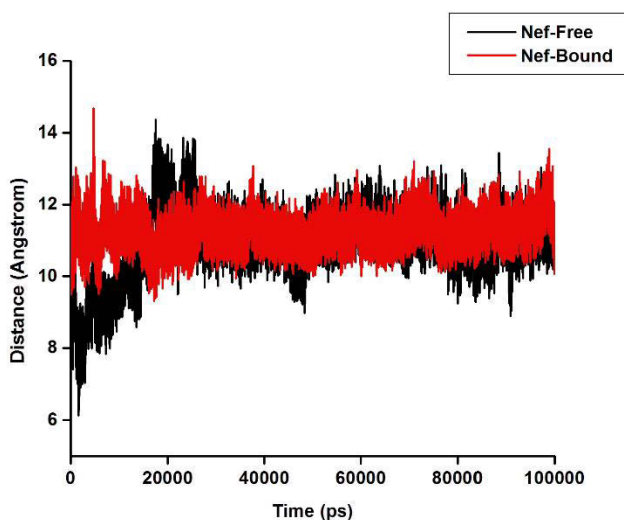

Figure 9.10. Distance between C-alpha residues involving Leu112 residues from both subunits. The average distance in case of apo conformation (10.93 Å) was lower as compared to bound conformation (11.24 Å).
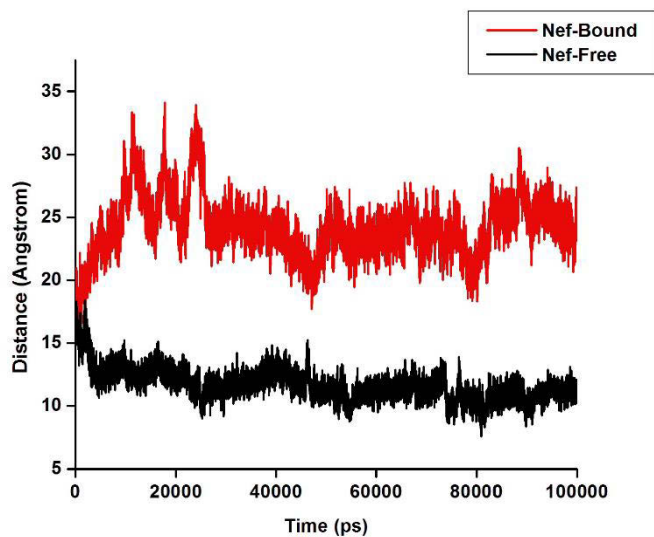
Figure 9.11. Distance between C-alpha residues involving Gln104's from both subunits. The average distance between two oppositely placed Gln104 residues found to be 24.63 Å and 11.62 Å for B9 bound and apo conformation of Nef respectively.
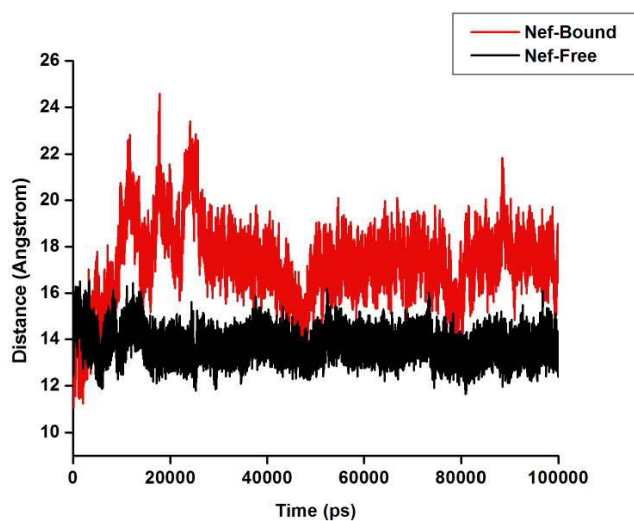

Figure 9.12. Distance between C-alpha residues involving Asp108 residues from both subunits. The average distances were 17.32 Å and 13.73 Å for bound and apo conformations of HIV-Nef respectively.
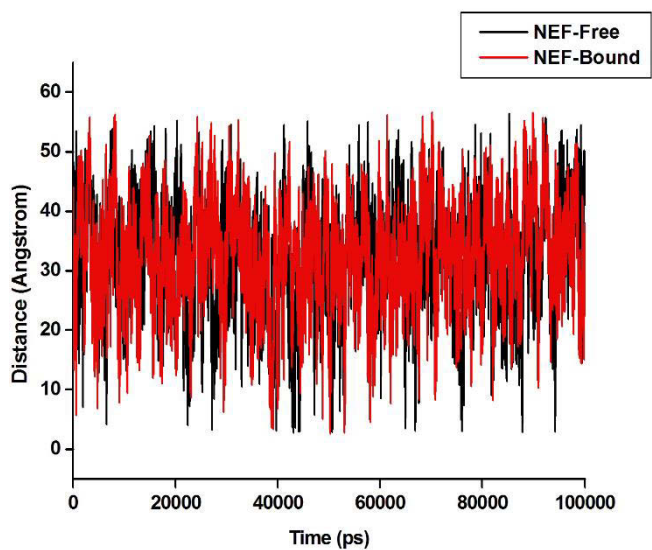
Figure 9.13. Distance between C-alpha residues involving Tyr115 residues from each monomer. The average distances were found to be 32.38 Å and 31.70 Å respectively for bound and apo conformations respectively.

Table 9.2. Residues involved in dimer packing [11] and their average distances from each other during simulation time.

Residues	Average Distance (Angstrom)
Asp108-Asp108	17.32* 13.73
Leu112-Leu112	11.24* 10.93
Gln104-Gln104	24.23* 11.62
Tyr115-Tyr115	32.38* 31.70

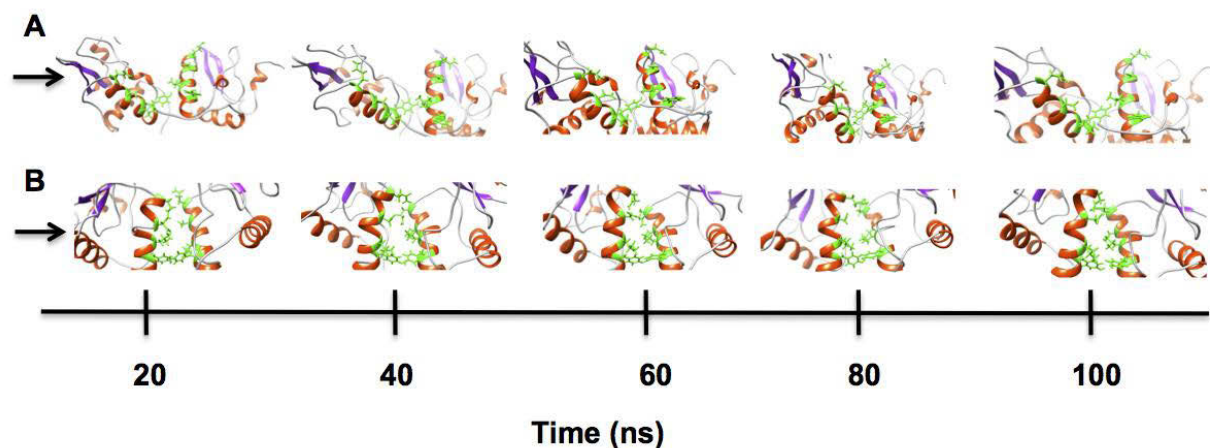*Denotes B9 bound conformation of HIV-Nef

Figure 9.14. Snapshots of apo and B9 bound conformations of HIV-Nef at a certain time interval during MD simulation. The residues highlighted in 'green' are responsible in the process of dimer packing. A and B highlights the pathway of dimer dissociation and association for inhibitor (B9) bound and apo conformation respectively.

Understanding the evolution of residues involved in dimer packing by analyzing the time dependent snapshots of apo and B9 bound conformation of Nef in combination with post dynamics analysis e.g. distance, Rg, RMSD, RMSF provides a model to understand the process of dimer packing and explain how the inhibitor gained access to the Nef dimeric site and its evolution in dimeric region leads to a dimeric inhibition. These outcomes from a classical molecular dynamics simulation unveils the large conformational drifts of HIV-Nef during the process of inhibitor binding which will prove effective in designing novel Nef dimerization inhibitors.

9.5.2.3. Dynamic Cross Correlation (DCC) analysis

Figure 9.15 highlights the cross correlation map calculated for apo and bound conformation of HIV-Nef. It is evident from the correlation map that more globally correlated motion is observed in case of the free conformation of HIV-Nef, which further justifies the fact that in free conformation HIV-Nef went through a process of dimerization causing a more correlated residue-residue interaction. On the contrary, in case of bound conformation, a greater existence of negative correlated motions during simulation time was observed. This fact is further justified by observing a higher occupancy of positive C_{ij} patches in case of free conformation as compared

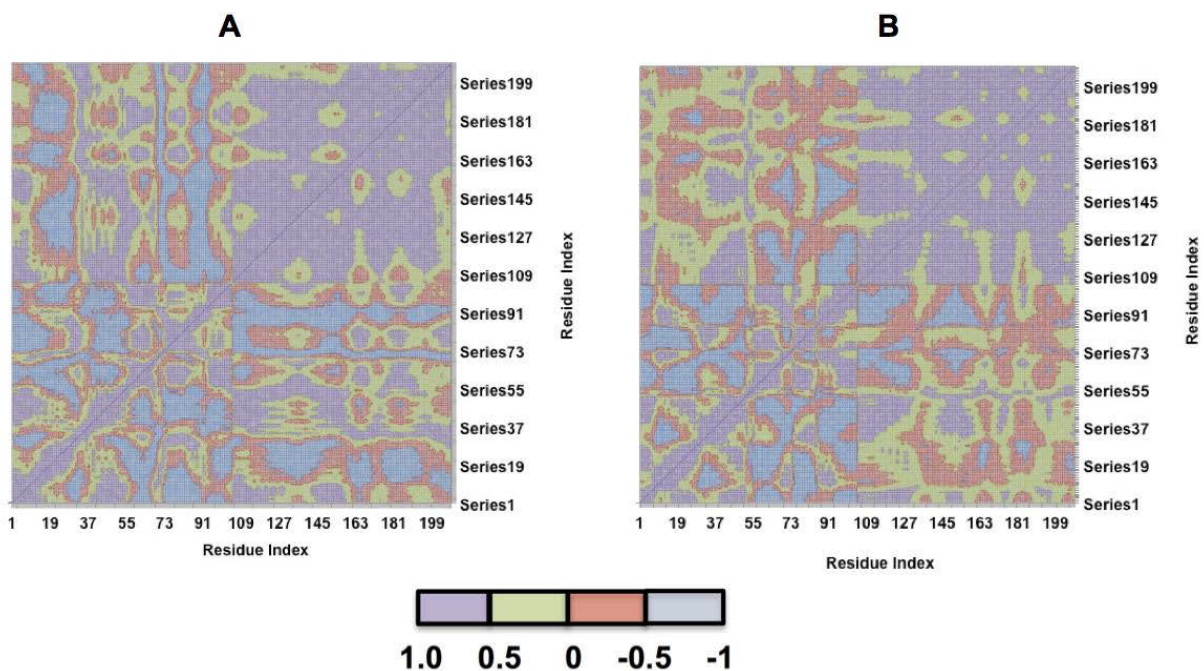to bound one. Such findings provide a solid conclusion on the mechanism of dimerization inhibition by B9.

Figure 9.15. DCC map during simulation time taking in account Cα residues of HIV-Nef ligand bound (A) and free (B) conformations.

9.5.2.4. Principal Component Analysis (PCA)

To further understand conformational preferences for both ligand-bound and unbound conformations of HIV-Nef, principal component analysis was carried out taking in account C-alpha residues of both the systems. **Figure 9.16** highlights the dominant changes in motion across two principal components in case ligand-bound and unbound configurations of HIV-Nef. It was found that eigenvectors computed from individual trajectories were quite varied between the two systems, which further emphasize on the difference in the conformational landscape between the free and ligand bound conformation. The difference in magnitude of Eigen values coming from first principal components of both apo and bound conformations found to be ~0.05 Å whereas this difference was ~0.02 Å in case of the second principal component. This difference in average Eigen values across first two principal components suggests a greater mobility of the apo conformation as compared to its inhibitor bound counterpart.
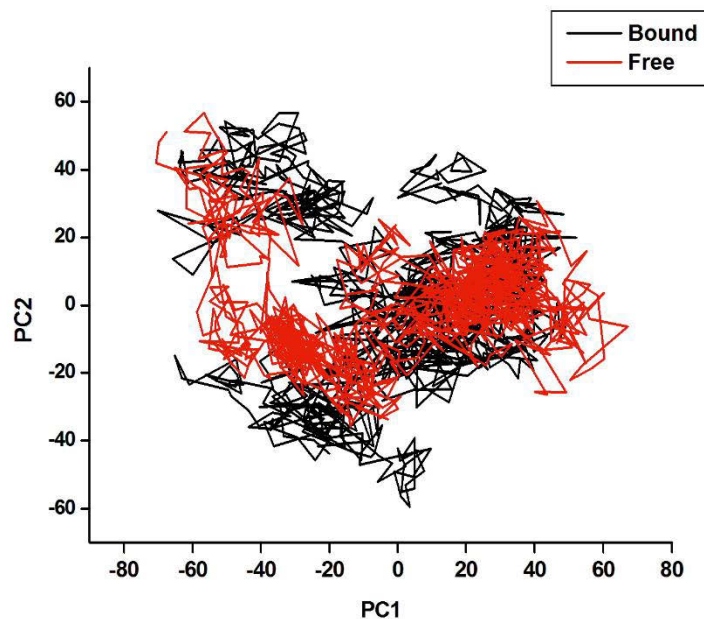
Figure 9.16. Projections of Eigen values during simulation period for ligand bound and apo (free) conformations of HIV-Nef along the first two principal components (PC1 and PC2).

Moreover porcupine plots across two normal modes clearly indicates a closer distance between helical dimers in free conformation when compared with B9 bound conformation of HIV-Nef (**Figure 9.17** and **9.18**). Also it clearly indicates an overall change in direction of motion between bound and unbound conformation of HIV-Nef.
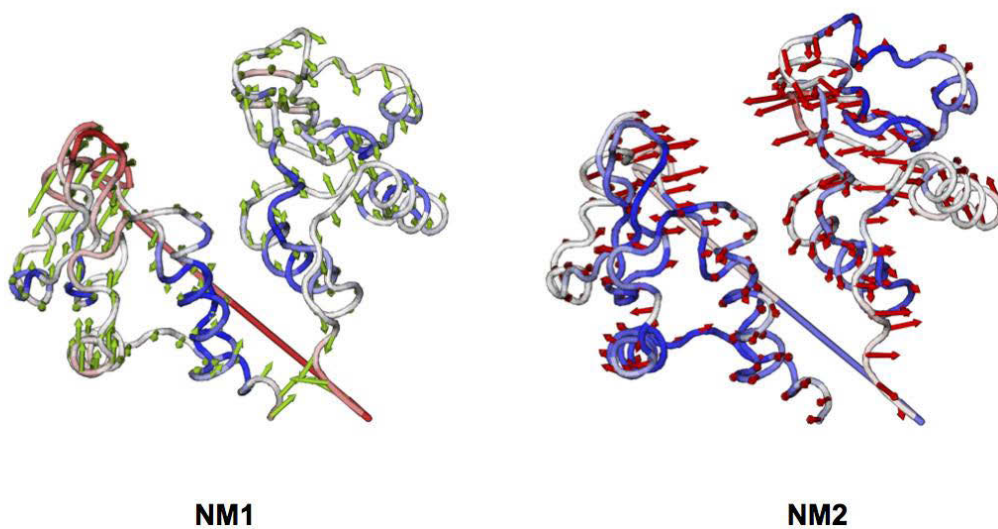

NM1

NM2

Figure 9.17. Porcupine plots across two different normal modes showing the direction of motion of unliganded (free) HIV-Nef system.
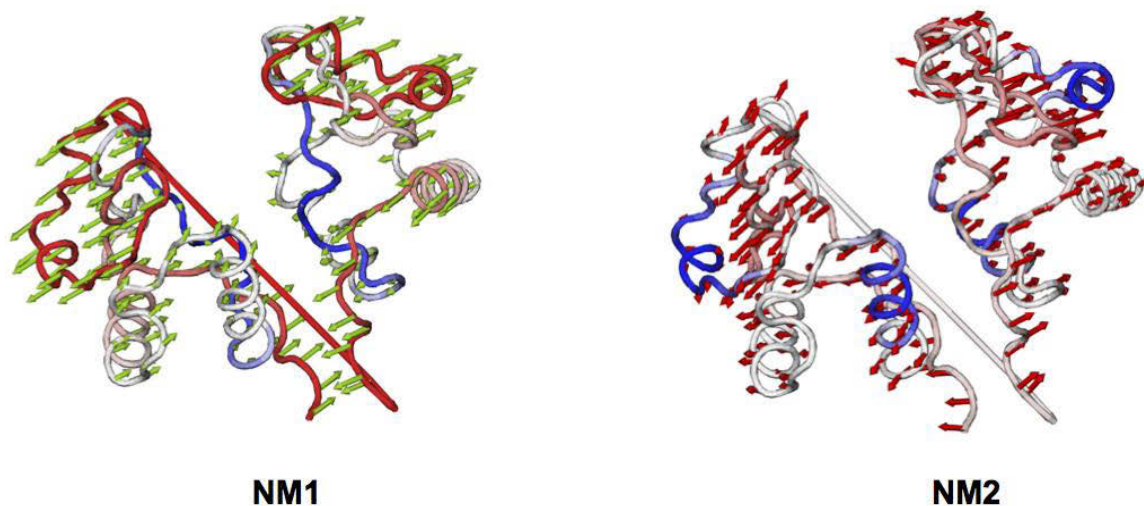
NM1

NM2

Figure 9.18. Porcupine plots across two different normal modes showing the direction of motion of ligand bound (B9) HIV-Nef system.

We also opted to monitor residue-based biomolecular flexibility across different normal modes in order to further understand the conformational flexibility and rigidity of the two systems, free and bound protein (**Figure 9.19** and **9.20.**).
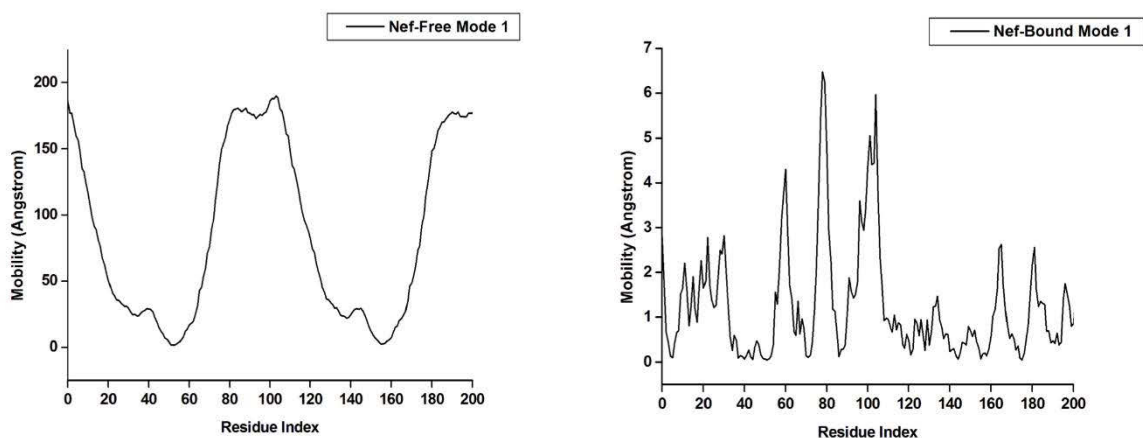

Figure 9.19. Comparison of mobility plot of unliganded and ligand bound conformation of HIV-Nef systems in normal mode 1.
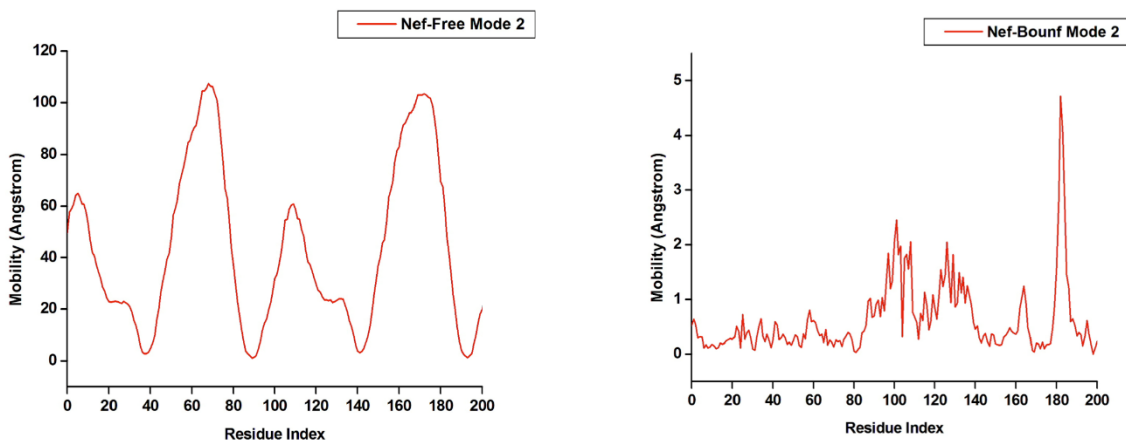
Figure 9.20. Comparison of mobility plot of unliganded and ligand bound conformation of HIV-Nef systems in normal mode 2.

Interestingly it was noticed that the flexibility of the residues ranging from 100 to 120 were comparatively higher in case of apo conformation of HIV-Nef across first two normal modes. The average difference in fluctuation between apo and B9 bound conformation of HIV-Nef within residues 100-120 found to be ~100 Å (mode 1) and ~5 Å (mode 2) respectively. This region contained residues involved in dimer packing. Normal mode analysis not only highlights the fact that the process of dimer packing affects the overall flexibility but also substantiates the findings of RMSD, Rg, and RMSF parameters and further justifies the process of dimerization and its impact on overall protein flexibility.

9.6. Conclusion

Molecular dynamics simulation reveals a dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations. Small molecule inhibitors such as B9, which targets the dimeric helical area of HIV-Nef, inhibits the process of dimerization thus leading to a more conformationally rigid system with hampered dimerization process. The RMSF, radius of gyration and mobility plots generated during normal mode analysis for both the systems suggested a more conformational flexible nature of HIV-Nef dimer in the absence of an inhibitor. The increased magnitude of parameters e.g. Rg (~1.5 Å), C-alpha deviations at the dimeric helix (~2 Å) suggested a greater conformational flexibility of Nef apo conformation and

a flexible dimeric helix. On the contrary, B9 bound conformation of HIV-Nef found to be more conformationally rigid with a lesser inter-dimeric association this the simulation period. Location of inhibitor, B9 in the active site of two helical subunits act as barrier in the process of dimerization which can be easily understood by monitoring the distance among residues involved in the process of dimerization and as well as by visual inspection of snapshots generated during the long range molecular dynamics simulation. The difference in magnitude of the distance parameter for inter-residue connection among Asp108, Leu112 and Gln104 found to be 4 Å, 1 Å and 12 Å respectively. This first account report highlights important dynamic features of an important HIV target, which would also serve as an initial point in the process of designing novel compounds against HIV-Nef as anti-HIV drugs.

9.7. Acknowledgements

9.8. Conflict of Interests

Authors declare no financial and intellectual conflict of interests.

9.9. References

1.	Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.

2.	Rambaut, A., Posada, D., Crandall, K. A., and Holmes, E. C. (2004) The causes and consequences of HIV evolution, *D - 100962779 5*, 52-61.

3.	Xu, Y., Liu, H., Niu, C. Y., Luo, C., Luo, X. M., Shen, J. H., Chen, K. X., and Jiang, H. L. (2004) Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-

yl)-butanes based on the structural modeling of human CCR5 receptor, *Bioorganic & Medicinal Chemistry 12*, 6193-6208.

4. Soliman, M. E. S. (2013) A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors, *Drug Development Research 74*, 283-295.

5. Johnson, B. C., Pauly, G.T., Rai, G., Patel, D., Bauman, J.D., Baker, H.L., Das, K., Schneider, J.P., Maloney, D.J., Arnold, E., Thomas, C.J., and Hughes, S.H. . (2012) A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants, *Retrovirology* 1-23.

6. Morah, E. U. (2007) Are People Aware of Their HIV-positive Status Responsible for Driving the Epidemic in SubSaharan Africa? The Case of Malawi In *Development Policy Review*, pp 215-242.

7. Bhakat, S., Martin, A. J. M., and Soliman, M. E. S. (2014) An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine, *Molecular BioSystems 10*, 2215-2228.

8. Sarafianos, S. G., Das, K., Hughes, S. H., and Arnold, E. (2004) Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases, *Current Opinion in Structural Biology 14*, 716-730.

9. Das, S. R., and Jameel, S. (2005) Biology of the HIV Nef protein, *Indian Journal of Medical Research 121*, 315-332.

10. Emert-Sedlak, L. A., Narute, P., Shu, S. T., Poe, J. A., Shi, H., Yanamala, N., Alvarado, J. J., Lazo, J. S., Yeh, J. I., Johnston, P. A., and Smithgall, T. E. (2013) Effector Kinase Coupling Enables High-Throughput Screens for Direct HIV-1 Nef Antagonists with Antiretroviral Activity, *Chemistry & Biology 20*, 82-91.

11. Breuer, S., Schievink, S. I., Schulte, A., Blankenfeldt, W., Fackler, O. T., and Geyer, M. (2011) Molecular Design, Functional Characterization and Structural Basis of a Protein Inhibitor Against the HIV-1 Pathogenicity Factor Nef, *Plos One 6*, e20033.

12. Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How Fast-Folding Proteins Fold, *Science 334*, 517-520.

13. Garcia, A. E. (2008) Molecular dynamics simulations of protein folding, *Methods in molecular biology (Clifton, N.J.) 413*, 315-330.

14. Gsponer, J., and Caflisch, A. (2002) Molecular dynamics simulations of protein folding from the transition state, *Proceedings of the National Academy of Sciences of the United States of America 99*, 6719-6724.

15. Swope, W. C., Pitera, J. W., Suits, F., Pitman, M., Eleftheriou, M., Fitch, B. G., Germain, R. S., Rayshubski, A., Ward, T. J. C., Zhestkov, Y., and Zhou, R. (2004) Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide, *Journal of Physical Chemistry B 108*, 6582-6594.

16. Durrant, J. D., and McCammon, J. A. (2011) Molecular dynamics simulations and drug discovery, *Bmc Biology 9*.

17. Wan, H., Hu, J.-p., Tian, X.-h., and Chang, S. (2013) Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d, *Physical Chemistry Chemical Physics 15*, 1241-1251.

18. Karubiu, W., Bhakat, S., and Soliman, M. S. (2014) Compensatory Role of Double Mutation N348I/M184V on Nevirapine Binding Landscape: Insight from Molecular Dynamics Simulation, *Protein Journal*, 1-15.

19. Borodin, O., and Smith, G. D. (2000) Molecular dynamics simulations of poly(ethylene oxide)/LiI melts. 2. Dynamic properties, *Macromolecules 33*, 2273-2283.

20. Falconi, M., Biocca, S., Novelli, G., and Desideri, A. (2007) Molecular dynamics simulation of human LOX-1 provides an explanation for the lack of OxLDL binding to the Trp150Ala mutant, *Bmc Structural Biology 7*.

21. Lee, C. H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996) Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain, *Cell 85*, 931-942.

22. 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.

23. Moonsamy, S., and Soliman, M. E. S. (2014) Dual acting HIV inhibitors: integrated rational in silico design strategy, *Medicinal Chemistry Research 23*, 682-689.

24. Moonsamy, S., and Soliman, M. E. S. (2014) Computer-Aided Perspective for the Design of Flexible HIV Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs): de-novo Drug Design, Virtual Screening and Molecular Dynamics Simulations, *Letters in Drug Design & Discovery 11*, 513-524.

25. Trott, O., and Olson, A. J. (2010) Software News and Update 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.

26. Lau, C. D., Levesque, M. J., Chien, S., Date, S., and Haga, J. H. (2010) ViewDock TDW: high-throughput visualization of virtual screening results, *Bioinformatics 26*, 1915-1917.

27. Goetz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, *Journal of Chemical Theory and Computation 8*, 1542-1555.

28. Ryckmans, T., Balancon, L., Berton, O., Genicot, C., Lamberty, Y., Lallemand, B., Pasau, P., Pirlot, N., Quere, L., and Talaga, P. (2002) First dual NK1 antagonists-serotonin reuptake inhibitors: Synthesis and SAR of a new class of potential antidepressants, *Bioorganic and Medicinal Chemistry Letters 12*, 261-264.

29. 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.

30. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins-Structure Function and Bioinformatics 78*, 1950-1958.

31. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparsion of simple potential functions for simulating liquid water, *Journal of Chemical Physics 79*, 926-935.

32. Harvey, M. J., and De Fabritiis, G. (2009) An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware, *Journal of Chemical Theory and Computation 5*, 2371-2377.

33. Roe, D. R., and Cheatham, T. E., III. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *Journal of Chemical Theory and Computation 9*, 3084-3095.

34. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics & Modelling 14*, 33-38.

35. Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: Protein Dynamics Inferred from Theory and Experiments, *Bioinformatics 27*, 1575-1577.

36. Iyer, P. C., Zhao, J., Emert-Sedlak, L. A., Moore, K. K., Smithgall, T. E., and Day, B. W. (2014) Synthesis and structure-activity analysis of diphenylpyrazolodiazene inhibitors of the HIV-1 Nef virulence factor, *Bioorganic & Medicinal Chemistry Letters 24*, 1702-1706.

37. Moonsamy, S., Bhakat, S., and Soliman, M. E. S. Identification, binding mode and prospective chemical structural features of novel Nef protein inhibitors as potential anti-HIV drugs, *Medicinal Chemistry Research*.

CHAPTER 10

10. General conclusions and future study recommendations

10.1. General conclusions

The six major aims of this study were to:

1. **Chapter 4 - To design novel inhibitors with potential dual activity against HIV Protease (PR) and Reverse transcriptase (RT) enzymes:**

   On the basis of structural information retained in known HIV-1 PR and RT inhibitors and understanding their respective binding mechanisms to specific targets, five novel leads were proposed as potential dual acting PR/RT inhibitors. Herewith, using a hybrid "loop docking"/MD strategy as well as per-residue interactions calculations, the five proposed structures exhibited better binding in comparison to prototype drugs. Amongst these novel leads, compound 2 with a bis-tetrahydrofuranyl moiety displayed the most inhibition activity against both PR and RT enzymes.

2. **Chapter 5 - To verify the activity of hydroquinone-based compounds as non-nucleoside reverse transcriptase inhibitors (NNRTIs), as proposed by Bruccoleri:**

   Previously, flexible hydroquinone-based compounds were proposed as potential mutant-resistant NNRT inhibitors, but no experimental or computational work supported this proposal. Therefore, *via* an integrated *in-silico* computational framework that included *de-novo* design, structure-based virtual screening (SBVS), MD simulations and post-dynamic per-residue binding energy decomposition analysis, the binding affinity as well as the interaction landscape, novel flexible hydroquinone-based compounds exhibited NNRT inhibitor activity, respectively. These proposed leads demonstrated improved binding affinity as compared to FDA-approved NNRTI prototype drugs (i.e. rilpvirine,

nevirapine and efavirenz), however, the bioavailability profile of these compounds could hamper their uses as effective drugs.

3. **Chapter 6 - To identify novel CCR5 antagonists as potential HIV-1 entry inhibitors using integrated computational tools:**

Using an integrated *in-silico* computational approach, which included homology modelling, hybrid structure-based and pharmacophore-based virtual screening, MD simulations, per-residue energy interactions calculations and atom-based 3D-QSAR analysis, ten potential compounds were proposed as novel CCR5 antagonists or potential HIV-1 entry inhibitors. These ten novel leads *via* validated docking binding free energy calculations revealed better binding affinities with CCR5 as compared to the only FDA-approved HIV-1 entry inhibitor, maraviroc. Per-residue energy interactions calculations on averaged MD structures demonstrated that hydrophobic active residues Trp86, Tyr89 and Tyr108 contributed the most to inhibitor binding. The validated atom-based 3D-QSAR model exhibited a high cross-validated r_{cv}^2 value of 0.84 using three principal components and a non-cross-validated r^2 value of 0.941. It was also observed that nearly all compounds in the test set and training set yielded a good predicted value.

4. **Chapter 7 - To provide a more comprehensive insight into the precise impact of the M184I mutation on RT resistance to lamivudine:**

The underlying atomistic nature surrounding the mystery of the M184I single point mutation that interferes and in turn leaves HIV-1 RT totally resistant to lamivudine was investigated and accomplished using MD simulations, binding free energy calculations, PCA and RIN analysis, respectively. Results revealed that single point mutations at residue 184 of RT caused; (1) distorted orientation of lamivudine within the active site as a result of steric confliction between the oxathiolane ring of lamivudine and side chain of beta-branched amino acids Ile at position 184 and therefore, perturbs inhibitor binding, (2) a reduction in the binding affinity (~8 kcal/mol) compared to the wild type protein, (3) disparity in overall protein motions evident from PCA performed on both systems and

(4) alteration of the hydrogen bonding framework and atomic interactions with the inhibitor.

5. **Chapter 8 - To identify more potent potential HIV-Nef inhibitors by exploiting the structural features of B9 using an integrated computational tools framework:**

To date, only one compound, known as B9, has been reported as a potent Nef inhibitor. Herewith, using a hybrid ligand shape-similarity-and pharmacophore-based virtual screening tactic, in conjugation with MD and further post-dynamic analysis, we identified and proposed top ten compounds from each compound library as new potential novel leads targeting HIV-Nef with a detailed analysis of their respective binding modes. Furthermore, the top identified hit compounds from the shape similarity-based library (**ZINC04177596**, Δ**G**$_{bind}$= -28.7482 kcal/mol**)** and pharmacophore-based library (**ZINC36617540**, Δ**G**$_{bind}$= -20.2271 kcal/mol**)** possessed comparatively better binding affinities compared to the reference molecule, B9 (Δ**G**$_{bind}$= -18.0694 kcal/mol). Also, both top identified hits; **ZINC04177596** and **ZINC36617540** demonstrated relatable binding modes, as the prototype antagonist B9, at the Nef active site. The prominent binding forces were hydrophobic and electrostatic interactions, which were responsible for holding these ligands tightly in place at the Nef dimer interface.

6. **Chapter 9 - To reveal the dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations:**

Molecular dynamic simulations revealed a dimer packing and unpacking phenomena of HIV-Nef in its apo and inhibitor bound conformations. The only known small molecule Nef inhibitor, B9, targeted the dimerc helical area of HIV-Nef and inhibited the dimerization process, and thus produced a more conformational rigid system with a hindered dimerization process. During normal mode analysis conducted for both protein systems, the generated RMSF, radius of gyration and mobility plots advocated a more conformational flexible nature of HIV-Nef dimer in the absence of the inhibitor. As mentioned earlier, the Nef inhibitor B9, located at the dimeric active site acts as a barrier

during the dimerization process, and thus is easily understood by observation of inter-residue distance parameters involved in dimerization and also by visual inspection of snapshots generated during the long range MD simulations. The disparity in magnitude of inter-residue connection distance parameters amongst Asp108, Leu112 and Gln104 were found to be 4 Å, 1 Å and 12 Å, respectively.

10.2. Future study recommendations

1. **Chapter 4:** Herein we reported the main concept and structural scaffolds of proposed dual acting PR/RT HIV inhibitors. In future, more meaningful binding free energy results could be achieved by using experimental data as a point of reference. Also extensive study on the relative binding affinities of more drugs against different PR and RT subtypes. The proposed structural scaffolds generated in this study could be used as a starting point for screening of a wider set of substituents for further lead optimization. Thus, synthesis and biological testing of proposed scaffolds is advised. Furthermore, newly developed structures and can be subjected for more computational analysis, for instance, the parameters used to calculate binding free energy needs to be optimized for more meaningful results. In order to provide more understanding into the proposed and other newly generated substituents dual activities and enzyme dynamics, advanced computational techniques such as PCA, RIN, QSAR and Substrate Envelope Analysis (SEA). Moreover, the *in-silico* strategy introduced in this study could be useful in the process of drug design and development, not only against HIV, but also for a wide range of other biological targets.

2. **Chapter 5:** *Via* an integrated *in-silico* computational framework, the results in this study attempted to provide a contribution towards verifying the activity of hydroquinone compounds initially proposed by Bruccoleri. However, the experimental aspect is lacking, therefore the proposed scaffolds from this study could be subjected to further synthesis and biological testing to provide experimental support to the above-mentioned proposal. These proposed leads demonstrated improved binding affinity as compared to FDA-approved NNRTI prototype drugs, but, the bioavailability profile of these

compounds could hamper their uses as effective drugs. Results obtained from this extensive study could assist medicinal and biochemistry researchers with further experimental investigations. Advanced computational techniques such as PCA, RIN, QSAR and Substrate Envelope Analysis (SEA) could be used to provide deeper insight into inhibitor activity and protein dynamics.

3. **Chapter 6:** The results in this study contributed towards shedding light on the activity of a new series of lead compounds as potential HIV entry inhibitors. Information gained from this study and the new series of novel lead compounds could serve as a powerful tool in the drug design and development process. Further screening and biological testing could be performed to provide validation to the computational work. Other advanced computational techniques such as PCA, RIN, SEA and QM/MM could be conducted to provide deeper insight into the different inhibitors activities and protein dynamics as well as protein-inhibitor interactions at the atomic level.

4. **Chapter 7:** The comprehensive analysis presented in this report could provide useful information in future for researchers for understanding the drug resistance mechanism against lamivudine. The results could also provide some potential clues for further drug design of novel inhibitors that are less susceptible to drug resistance and contribute to the development of anti-HIV thereapy. The findings of this report could provide potential markers for further synthetic and biological testing of more novel drugs that are less susceptible to drug resistance.

5. **Chapter 8:** The results in this study contributed towards the discovery of more potent inhibitors targeting HIV-Nef through exploitation of the structural features of B9 by using an integrated computational tools approach. As a result of the significant difference in the binding affinities between the identified hit compounds, ZINC36617540 and ZINC04177596 and the reference prototype drug, this prompted us to propose a set of structural criteria that could be imperative for further lead optimization and design of more potential small molecules targeting HIV-Nef. Further experimental investigation i.e.

screening and biological testing by medicinal and biochemists would boost the proficiency of the current report. Other advanced computational tools such as PCA, RIN, SEA, DCC and QSAR could be utilized to provide deeper understanding into inhibitor activity and protein dynamics as well as protein-inhibitor interactions at the atomic level.

6. **Chapter 9:** This first account report highlights important dynamic features of an important HIV target, which would also serve as an initial point in the process of designing novel compounds against HIV-Nef as anti-HIV drugs. The validation and proficiency of this current report could be achieved by further experimental investigation. Also, potential Nef inhibitors could be designed that focused solely on targeting the dimeric helical site of the HIV-Nef protein and possibly by using the QSAR methodology and then subjecting this series of compounds to further biological testing as a point of reference.

APPENDIX: PUBLISHED

Appendix A: Pdf version of the publication

Moonsamy, S. and Soliman, M.E. (2014) Dual Acting HIV Inhibitors: Integrated Rational *in-silico* Design Strategy, *Medicinal Chemistry Research*, 23, 682-689. (***Published***)

Appendix B: Pdf version of the publication

Moonsamy, S. and Soliman, M.E. (2014) Computer-aided perspective for the design of flexible HIV non-nucleoside reverse transcriptase inhibitors (NNRTIs): de-novo design, virtual screening and molecular dynamics simulations, *Letters in Drug Design and Discovery*, 11, 513-524. (***Published***)

Appendix C: Pdf version of the publication

Moonsamy, S., Dash, R.C., and Soliman, M.E. (2014) Integrated computational tool for identification of CCR5 antagonists as potential HIV-1 entry inhibitors: homologymodeling, virtual screening, molecular dynamics simulations and 3D QSAR analysis, *Molecules,* 19, 5243-5265. *(Published)*

Dual acting HIV inhibitors: integrated rational in silico design strategy

Suri Moonsamy & Mahmoud E. S. Soliman

ONLINE FIRST

Springer

Springer

**MEDICINAL
CHEMISTRY
RESEARCH**

ORIGINAL RESEARCH

Dual acting HIV inhibitors: integrated rational in silico design strategy

Suri Moonsamy · Mahmoud E. S. Soliman

Abstract In silico pharmacophore and structure-based drug design guided by binding mode analysis and molecular dynamics simulations is an effective approach for identifying novel potent and selective inhibitors. Herein, we demonstrated a unique strategy for developing dual acting inhibitors against HIV-1 protease and reverse transcriptase via an integrated computational protocol that relies on "loop docking" and molecular dynamics simulations, the designed targets exhibited binding affinities comparable to, and in some cases better than, known active reference drugs.

Introduction

Three main HIV-1 enzymes have been exploited as the core of chemotherapy for AIDS treatment. These enzymes are HIV-1 protease (PR), reverse transcriptase (RT), and integrase (IN) (Pani *et al.*, 2002). However, designing inhibitors to target each single enzyme have proved a successful strategy for treatment, the therapeutic effect of these inhibitors is drastically hampered by many implications. Nucleoside RT inhibitors (NRTIs) exhibited high-

S. Moonsamy · M. E. S. Soliman (✉)
School of Health Sciences, University of KwaZulu-Natal,
Durban 4001, South Africa
e-mail: soliman@ukzn.ac.za

intrinsic toxicity (Schinazi *et al.*, 2006), whereas, the less toxic protease inhibitors (PIs), non-nucleoside RT inhibitors (NNRTIs), and integrase inhibitors (INIs) are severely compromised by the rapid emergence of resistant viral strains (Young and Kuritzkes, 2002; Bacheler, 1999; Mbisa *et al.*, 2011).

A common tactic to overcome HIV disease resistant to a single therapy is the use of multidrug therapy. Highly active antiretroviral therapy (HAART) (Yeni, 2006) combines NRTIs with NNRTIs or PIs and effectively reduces HIV viral load which in turn considerably improves the life quality of AIDS patients (Palella *et al.*, 1998). Even though, HARRT proved successful, to ensure the efficacy of this primary treatment for AIDS, complete adherence—which is very difficult to attain because of complicated dosing and intolerable acute and chronic toxicities—is essential. Failure to achieve adherence typically results in viral rebound and, even worse, multidrug resistance.

To this end, we envisage a dual acting anti-HIV therapy using a single drug that concomitantly inhibits two viral enzymes that could result in lower toxicity, simplified dosing, and improved patient adherence, therefore reducing the possibility of drug resistance.

The concept of designing of multifunctional ligands has been proposed before (Morphy and Rankovic, 2009). Multifunctional ligands were classified into three categories on the basis of how the scaffolds were combined (Morphy and Rankovic, 2009): (a) conjugate, if scaffolds are linked by a linker, (b) fused, if they directly coupled with no linker, and (c) merged, if the scaffolds are integrated into a single structure sharing common feature. Another classification based on the targeted protein binding sites was also introduced. Targeted protein binding sites could be: (a′) adjacent within a single protein structure (Lewis *et al.*, 2002; Munoz-Torrero and Camps, 2006), (b′)

located in different proteins but recognize similar endogenous ligands (Gemma *et al.*, 2006; Yamamoto *et al.*, 2002), and (c′) located in different proteins but recognize unrelated ligands (Kawanishi *et al.*, 1996; Ryckmans *et al.*, 2002; Groneberg *et al.*, 1999). In this report, we propose potential inhibitors based on the above-mentioned (c and c′) classifications.

In this study, we aimed to design novel inhibitors with potential dual activity against HIV PR and RT enzymes. Our design is initially based on the pharmacophoric features conserved in most potent FDA-approved PR and NRT inhibitors (Fig. 1). Then, the pharmacophore-based created scaffolds were subjected to structure-based screening against the respective enzyme to estimate their binding affinities.

In the current work, the strategy to design potential dual acting inhibitors is assisted by an integrated in silico approach that exploits the structural features of both the inhibitor and the enzyme active site. With no doubt that the most appropriate way to verify the activity of in silico designed target is the experimental in vitro and/or in vivo testing, however, the time and cost factors should be considered carefully before testing anticipated inhibitors. Validated computational tools could serve as useful tools to save time and effort in the drug design process. In this report, we introduce a "tough" in silico approach that relies on the use of docking calculations followed by molecular dynamics simulations to accurately estimate the binding affinity, as well as the stability of the inhibitor–enzyme complexes. As a matter of fact, we believe that docking calculations could be entirely misleading and even using the same software, in many case, may result in different predictions. In order to ensure reliable docking results, we embarked on a novel approach which we name "loop-docking" to enhance the docking calculations predications (see Methods: in silico design and validation
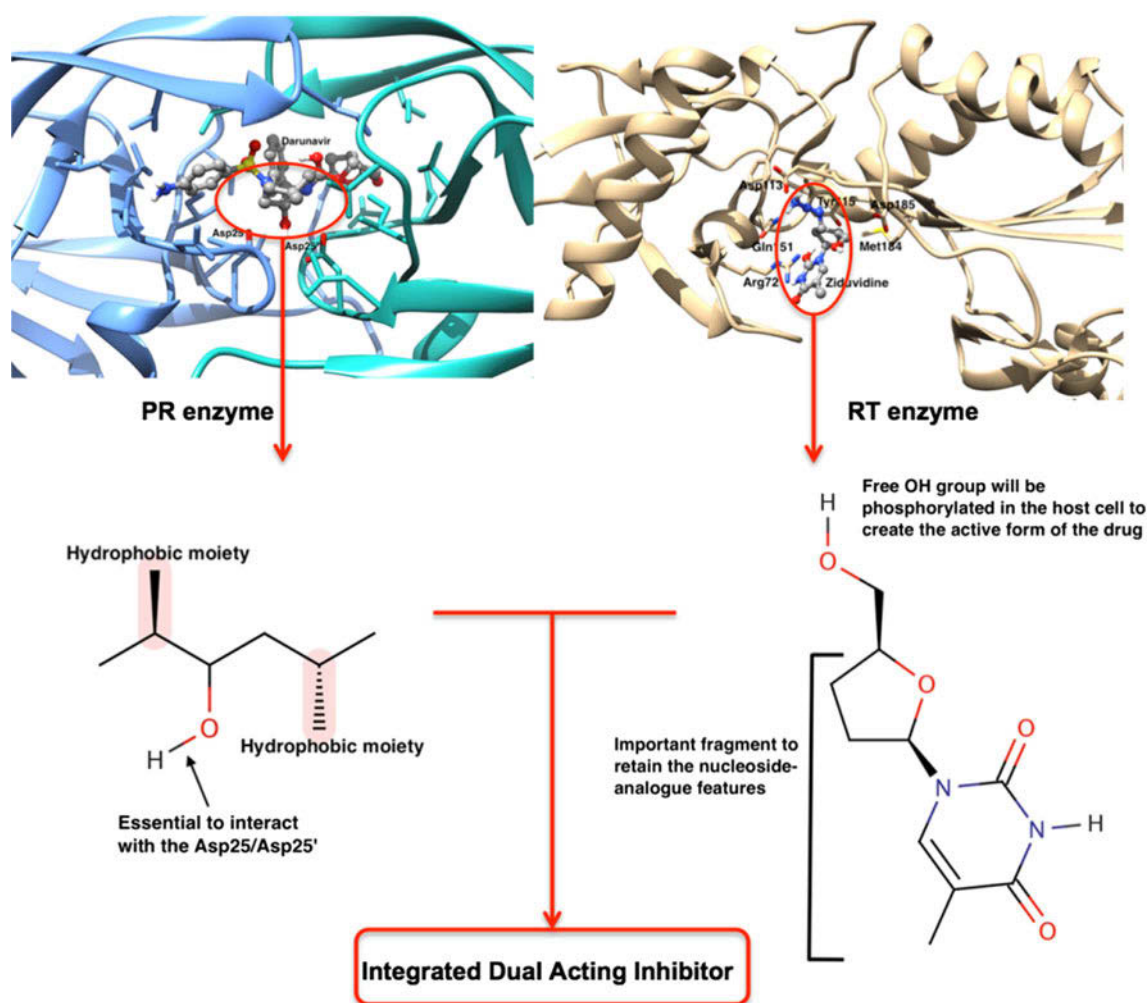

Fig. 1 The design of dual acting inhibitors is based on the pharmacophoric features of known inhibitors for each enzyme target and the binding theme of these pharmacophoric groups with the respective enzyme active site. Darunavir–PR complex (PDB code: 3QOZ) and Zidovudine–RT complex (PDB code: 3KLF) were used in this study to assign the pharmacophoric moieties

sections). Furthermore, the docked structures were then subjected to molecular dynamics simulations to ensure that these docked structures are stable enough over a reasonable time scale. Docked structures that fail to exhibit stability during MD simulations are subjected to further docking or rejected. The "loop docking"/MD strategy has been tested and proved successful on 30 X-ray crystal structures of various ligand–protein complexes (a manuscript that fully describe this approach has been submitted and under review). Herewith, we report the main concept and structural scaffolds of proposed dual acting PR/RT HIV inhibitors. Screening of a wider set of substituents for further lead optimization is still running. Synthesis of these compounds is quite feasible and most starting structures are commercially available. Since the synthesis and testing of the proposed inhibitors is still on-going, detailed information about the synthetic procedure, structural analysis, and biological testing will be revealed in a forthcoming publication. Previously, we have released the X-ray crystal structure of the HIV PR subtype C (Ahmed *et al.* 2013), as well as several reports on the synthesis, NMR and biological testing of a series of anti-HIV PR inhibitors (Honarparvar *et al.*, 2012; Karpoormath *et al.*, 2012; Makatini *et al.*, 2011a, b, 2013). Also, we demonstrated an extensive study on the relative binding affinities of nine anti-HIV FDA-approved drugs against different PRs subtypes (Ahmed *et al.* 2013).

The unique strategy adopted in this report could serve as a powerful tool in the process of drug design and development, not only against HIV, but also for a wide range of biological targets as well.

Methods: in silico design and validation

Constructing of the proposed structures and conformational ensembles

Proposed structures were constructed and energy minimized using Avogadros software (Hanwell *et al.*, 2012). To ensure thorough conformational sampling of the ligand candidates, low-energy conformations were generated for each molecule using Omega v1.8b (Hawkins *et al.*, 2010). For each compound, conformers generated within a range of 3.0 kcal/mol from the lowest energy conformer were investigated.

Protein systems

Protein structures were obtained from the protein data bank (http://www.rcsb.org/). The PDB codes 3QOZ and 3KLF were used for PR and RT, respectively. Procedure for enzyme model preparation for docking and MD simulations

is explained in details elsewhere (Makatini *et al.*, 2011a; Karpoormath *et al.*, 2012; Honarparvar *et al.*, 2012).

Docking calculations: "loop docking"

The idea behind our in-house "loop docking" approach is based on the fact that consecutive docking runs could remarkably improve the docking energy and orientation. From our experience with docking, we realized that in many cases the best-docked structure could be a docking artifact and does not represent the best docking orientation. Therefore, we always opt to rerun the docking calculation using the best-docked structure from initial docking as a starting structure for a second docking run and so on. We used a few scripts to allow this process to be automated. This automated "loop docking" will continue until a threshold value (δ) is reached. The threshold value (δ) is the difference between the docking binding energy of the last run and the preceding one. A threshold value of 0.05 was found to be appropriate. When this imposed value reached, the docking stops and the best-docked structure is selected. Autodock software (Morris *et al.*, 1998) is used for docking calculations. Details on docking set up and parameters can be found in our previous reports (Makatini *et al.*, 2011a; Karpoormath *et al.*, 2012; Honarparvar *et al.*, 2012).
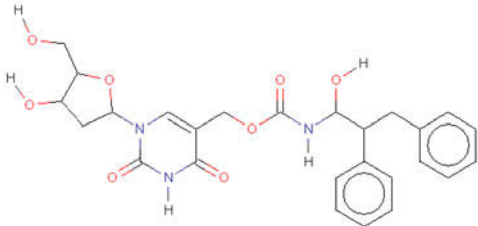
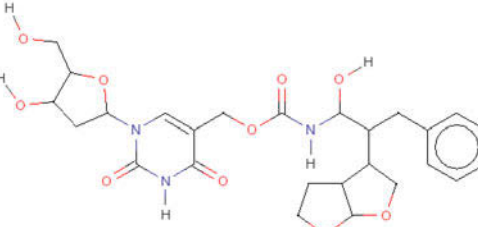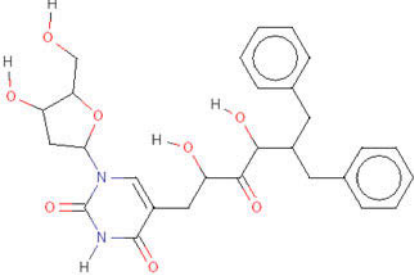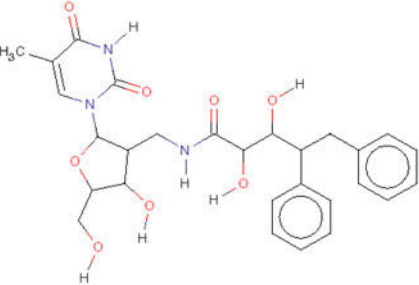### Molecular dynamics (MD) simulations

As before (Ahmed *et al.* 2013), MD simulations were performed for all inhibitor-enzyme complexes (eight systems in total) using Amber software (Case *et al.*, 2005).

Results and discussion

Design plan: pharmacophore and structure-based design

Our design began with the chemical knowledge of the structural features of PR and RT inhibitors, as well as the anatomy of the binding sites of both enzymes. First, we carefully studied the binding mode for all up-to-date resolved HIV PR and RT X-ray crystal structures of ligand–enzyme complexes to determine the common and most conserved pharmacophoric groups that critically involved in the ligand binding. However, finding a specific feasible way to help us easily analyze these ligand–enzyme binding themes was not an easy task to accomplish. After many trials, we came up with an efficient and viable procedure. For each enzyme class, PR and RT, all X-ray crystal structures were aligned using Chimera software (Pettersen *et al.*, 2004). This enabled us to easily focus only on the bound

Table 1 2D structures, the calculated binding energy, and chemical properties based on Lipinski's rule of five

Compound	2D structure	Calculated binding energy (kcal/mol)	
		PR DRV $(-9.4)^a$	RT AZT $(-6.2)^a$
1	 MWt (511.2), LogP(1.4), Ha (7), Hd(5)	−9.3	−7.1
2	 MWt (547.2), LogP(−0.13), Ha (9), Hd(5)	−9.5	−8.3
3	 MWt (524.5), LogP(1.7), Ha (8), Hd(5)	−8.5	−6.7
4	 MWt (539.5), LogP(0.50), Ha (9), Hd(6)	−8.4	−6.5

[a] Calculated binding energy for DRV and AZT against PR and RT enzymes, respectively

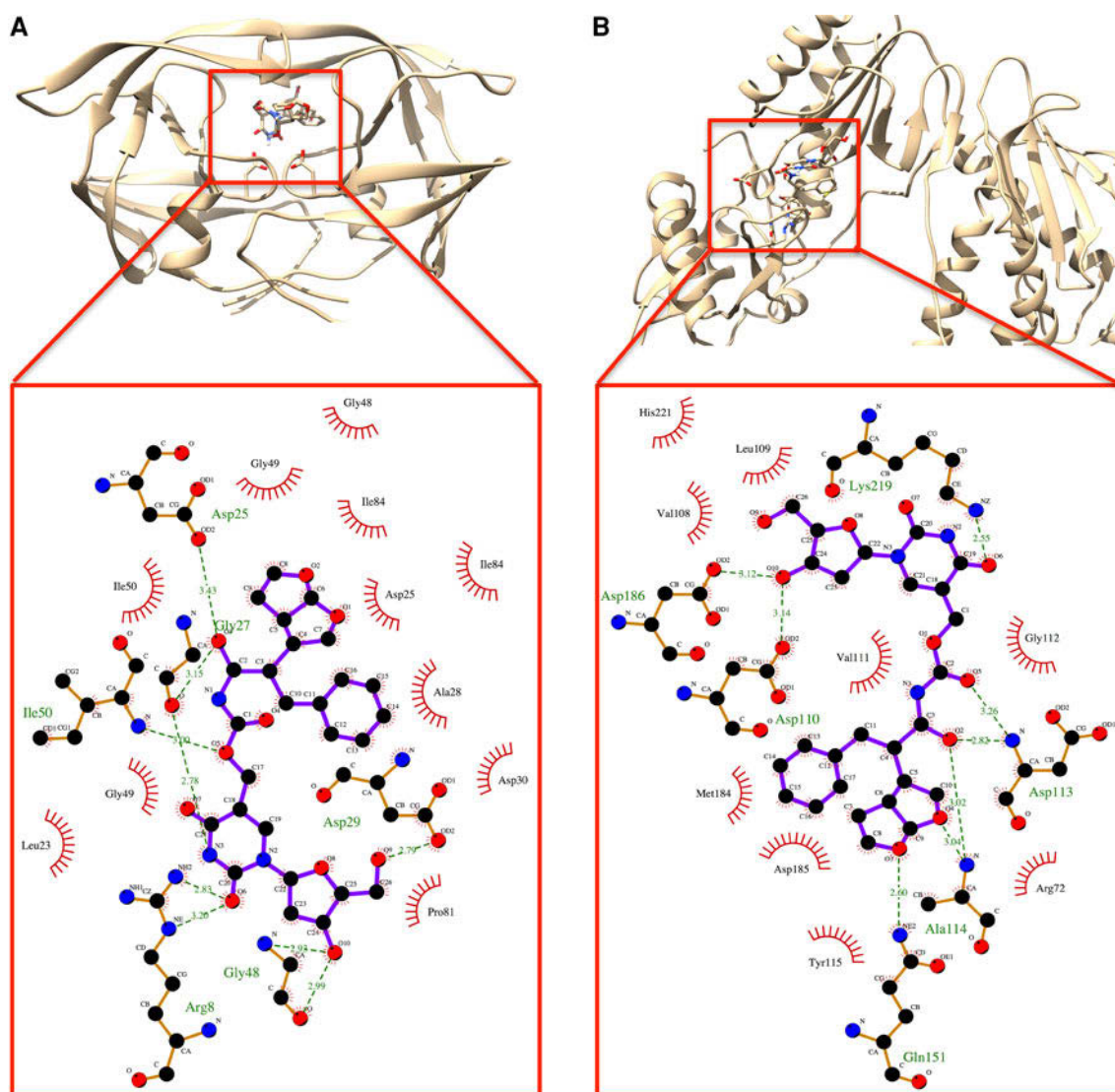[b] Calculated using MarvinSketch (http://www.chemaxon.com)

Fig. 2 Compound **2** in complex with PR and RT, **a** and **b**, respectively, showing the hydrogen bonding and electrostatic interactions with the enzymes active site

ligands to determine the common structural and binding components. Then, Ligplot software (Wallace *et al.*, 1995) was used to create a visual map of the most critical binding interactions between the ligands and enzyme active site. Information gained from this extensive structural and binding analysis allowed us to define the most important fragments for each class of inhibitors, PRIs and NRTIs (see Fig. 1). For PRIs, a characteristic structural feature is that hydrogen bond donors in the core of the ligand is important to form a hydrogen bond with the active site Asp25/Asp25′ residues. It has been shown that the interaction between the two catalytic aspartates and the hydroxyl group is worth more than 4 kcal mol (Appelt, 1993). Another feature is that the PR enzyme S1/S1′ subsites are mostly hydrophobic and ligand with hydrophobic groups that can fill in theses subsites have showed significant enzyme inhibitory activity

(Brik and Wong, 2003). Some studies have also reported that inhibitors that interact with the flap residues could be potential inhibitors (Damm *et al.*, 2008).

The same previous procedure has been repeated for the X-crystal structures of RT–NRTIs complexes to understand the binding theme of these inhibitors and obtain information on the most critical pharmacophoric moieties. Unlike PR inhibitors, it turned out that the binding behavior of NRTIs slightly varies from one crystal structure to another, however, seven amino acid residues seemed to be mainly involved in the binding with the inhibitors. These amino acids are Arg72, Gln151, Met184, Ala114, Asp113, Tyr115, and Phe116.

Based on all the above-mentioned compiled information, we proposed four main structural scaffolds (see Table 1) that we believe it could exhibit dual acting HIV
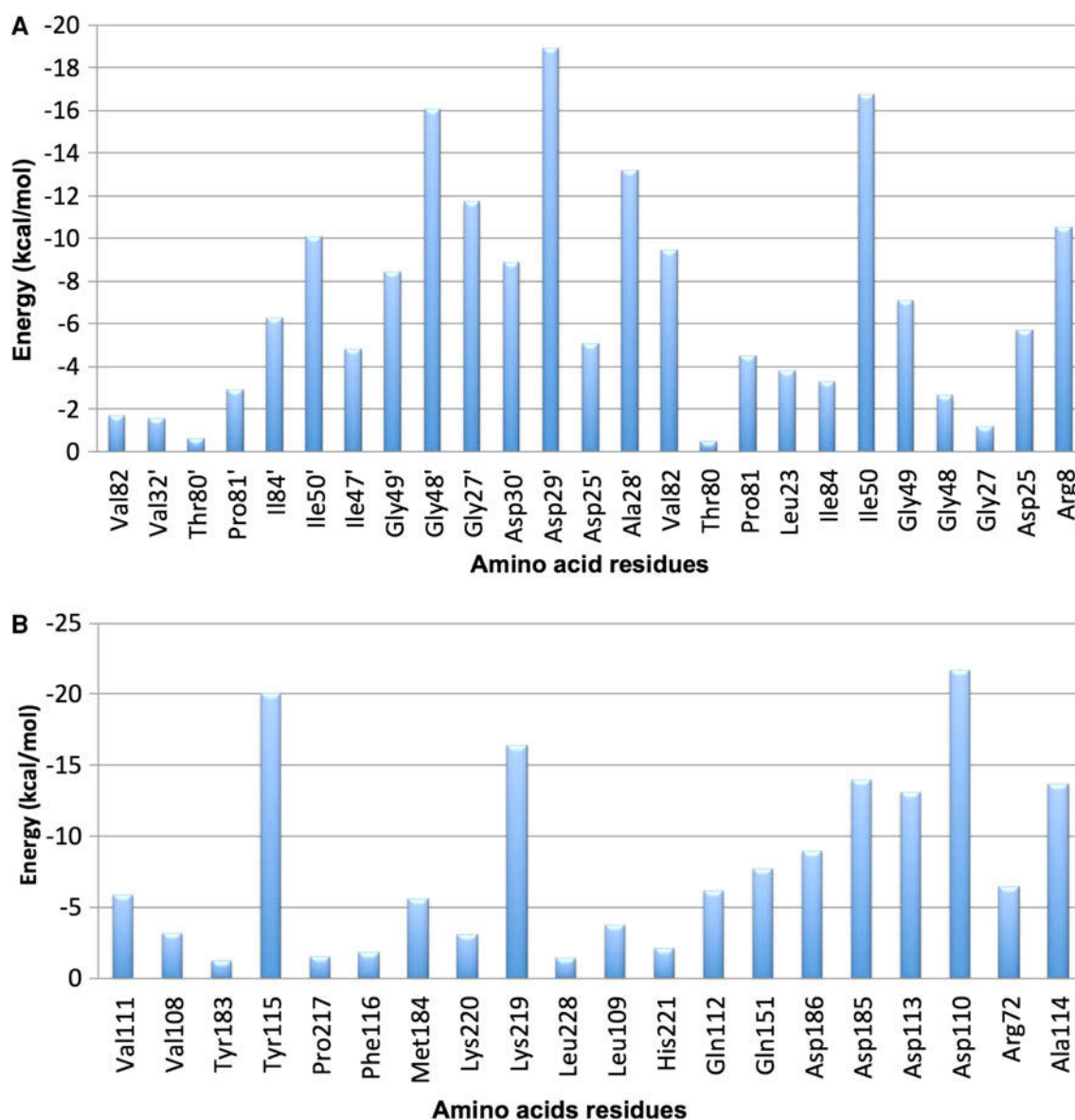
Fig. 3 Per-residue interactions of compound **2** with PR and RT, **a** and **b**, respectively

PR/RT inhibitory activity. Lead optimization and expansion phases, in which a larger set of compounds will be evaluated, are on-going. As mentioned above, synthesis of these proposed targets is quite feasible and almost all starting structures are commercially available.

Docking and MD simulations

The docking binding free energies for the proposed structures (**1-4**) against PR and RT are shown in Table 1. Docking of reference drugs, Darunavir and Zidovudine, was also performed for comparison (Table 1).

The docked complexes were then subjected to MD simulation for 5 ns to ensure the stability of the ligands in the enzyme active sites. Form our previous experience with

docking, in many cases, we found that, even the best docked structure flies away from the enzyme active site within a few picoseconds of MD simulations. Therefore, we believe that docking calculations that are not validated by relatively a long MD run to ensure the stability of the system is not reliable. Interestingly, for all the proposed compounds–enzyme complexes, the average RMSD values were below 2.3 Å and the variation of the potential energies falls within 1,000 kcal/mol and this is a good indication of the system stability. The PDB coordinates for all docked complexes (eight in total) are provided with the supplementary material.

As evident from Table 1, interestingly, structure **2** shows the best binding affinity toward PR and RT. This might due to the inclusion the bis-tetrahydrofuranyl

moiety. The bis-tetrahydrofuranyl moiety is found in Darunavir structure and as we anticipated, inclusion of this moiety resulted in a structure (compound **2**) with high-binding affinity comparable with Darunavir. As also seen from Table 1, all the proposed structures show better binding affinity toward RT than Zidovudine. This observation might be explained that inclusion of hydrophobic fragments in the structures increases hydrophobic interactions with the enzyme pocket residues and hence, improved the overall inhibitor binding.

Ligand–enzyme interactions

Post-dynamic complex structures were analyzed to ensure that the proposed structures fit well in the enzyme active pocket and, to some extent, exhibit similar interaction trends as the reference drugs with their respective enzyme targets. As evident from the hydrogen bonding and electrostatic interactions plots (Fig. 2), all proposed compounds showed, to a great extent, comparable interactions as in the crystal structures of the prototypes, DRV and AZT (all PDB coordinates are provided in the electronic supplementary material).

Per-residue interaction

For further lead optimization and expansion, it is useful to have a quantitative estimation of the contribution of each amino acid residues in the enzyme active site toward ligand binding. To accomplish this, we computed the per-residue interaction MolDock scoring function (Thomsen and Christensen, 2006). Insight obtained from per-residue interactions will allow us to tailor inhibitors that can strongly interact with those residues showed the most contribution toward the binding. Per-residue interactions for compounds **2** with PR and RT are shown in Fig. 3a and b.

From the hydrogen bonding interaction plots and per-residue interaction calculations, inhibitors clearly exhibited strong interaction with PR enzyme via hydrogen bond with the active site residues Asp25/Asp25′, a conserved interaction seen in all X-ray crystal structures of PIs-enzyme complexes. Interestingly, compound **2** showed a strong interaction with the flap residues, Ile50/Ile50′. Inhibitors that can interact with flap residues are believed to interfere with the flap opening and closure process, hence enzyme normal function (Damm *et al.*, 2008). With respect to RT, proposed structures showed hydrogen bonding as well as hydrophobic interactions with active site pocket residues, most importantly with Tyr115, Asp110, Arg72, and Ly219 (Figs. 2, 3). These interactions were also seen in the X-ray crystal structure of the AZT–RT complex.

As evident from the docking, MD simulations and per-residue calculations, the four proposed lead compounds are expected to be potential dual acting inhibitors and good starting point for further optimization and expansion.

The design strategy presented here could be implemented in the drug design and development process not only against HIV targets but also against a wide range of biological systems.

Conclusion

Based on the structural information of known HIV PR and RT inhibitors and the understanding of their binding mechanisms to the targets, novel leads were proposed as potential dual acting PR/RT inhibitors. Using hybrid computational approaches involving "loop docking" and molecular dynamics simulations and per-residue interactions calculations, the proposed structures were found to exhibit better binding affinities when compared to the prototype drugs. Compound **2** with bis-tetrahydrofuranyl moiety was found to be the most active compound against both PR and RT. The structures presented in this work are proposed as promising leads for developing potential drugs against HIV/AIDS therapy. Moreover, the in silico strategy introduced in this study could be useful in the process of drug design and development.

References

Ahmed SM, Kruger HG, Govender T, Maguire GE, Sayed Y, Ibrahim MA, Naicker P, Soliman ME (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

Appelt K (1993) Crystal structures of HIV-1 protease-inhibitor complexes. Perspect Drug Discov Des 1(1):23–48. doi:10.1007/bf02171654

Bacheler LT (1999) Resistance to non-nucleoside inhibitors of HIV-I reverse transcriptase. Drug Resist Update 2(1):56–67

Brik A, Wong C (2003) HIV-1 protease: mechanism and drug discovery. Org Biomol Chem 1:5–14

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688. doi:10.1002/jcc.20290

Damm KL, Ung PM, Quintero JJ, Gestwicki JE, Carlson HA (2008) A poke in the eye: inhibiting HIV-1 protease through its flap-recognition pocket. Biopolymers 89:643–652

Gemma S, Gabellieri E, Huleatt P, Fattorusso C, Borriello M, Catalanotti B, Butini S, De Angelis M, Novellino E, Nacci V, Belinskaya T, Saxena A, Campiani G (2006) Discovery of huperzine A-tacrine hybrids as potent inhibitors of human cholinesterases targeting their midgorge recognition sites. J Med Chem 49(11):3421–3425

Groneberg RD, Burns CJ, Morrissette MM, Ullrich JW, Morris RL, Darnbrough S, Djuric SW, Condon SM, McGeehan GM, Labaudiniere R, Neuenschwander K, Scotese AC, Kline JA (1999) Dual inhibition of phosphodiesterase 4 and matrix metalloproteinases by an (arylsulfonyl)hydroxamic acid template. J Med Chem 42(4):541–544

Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4:1–17

Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. J Chem Inf Model 50:572–584

Honarparvar B, Makatini MM, Pawar SA, Petzold K, Soliman ME, Arvidsson PI, Sayed Y, Govender T, Maguire GE, Kruger HG (2012) Pentacycloundecane-diol-based HIV-1 protease inhibitors: biological screening, 2D NMR, and molecular simulation studies. ChemMedChem 7(6):1009–1019

Karpoormath R, Sayed Y, Govender P, Govender T, Kruger HG, Soliman ME, Maguire GE (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(1):19–29. doi:10.1016/j.bioorg.2011.08.002

Kawanishi Y, Ishihara S, Tsushima T, Seno K, Miyagoshi M, Hagishita S, Ishikawa M, Shima N, Shimamura M, Ishihara Y (1996) Synthesis and pharmacological evaluation of highly potent dual histamine H-2 and gastrin receptor antagonists. Bioorg Med Chem Lett 6(13):1427–1430

Lewis WG, Green LG, Grynszpan F, Radic Z, Carlier PR, Taylor P, Finn MG, Sharpless KB (2002) Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. Angew Chem Int Ed Engl 41(6):1053–1057

Makatini M, Petzold K, Sriharsha NS, Ndlovu N, Soliman MES, Honarparvar B, Parboosing R, Naidoo A, Arvidsson PI, Sayed Y, Govender P, Maguire GEM, Kruger HG, Govender T (2011a) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D-NMR and docking/QM/MM/MD approach. Eur J Med Chem 46(9):3976–3985

Makatini M, Petzold K, Sriharsha S, Soliman MES, Honarparvar B, Arvidson PI, Sayed Y, Govender P, Maguire GEM, Kruger H, Govender T (2011b) Pentacycloundecane-based Inhibitors of wild-type C-South African HIV protease. Bioorg Med Chem Lett 15(21):2274–2277. doi:10.1016/j.bmcl.2011.02.105

Makatini MM, Petzold K, Alves CN, Arvidsson PI, Honarparvar B, Govender P, Govender T, Kruger HG, Sayed Y, Jeronimolameira, Maguire GE, Soliman MES (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors. J Enzym Inhib Med Chem 28(1):78–88. doi:10.3109/14756366.2011.633907

Mbisa JL, Martin SA, Cane PA (2011) Patterns of resistance development with integrase inhibitors in HIV. Infect drug resist 4:65–76

Morphy R, Rankovic Z (2009) Designing multiple ligands—medicinal chemistry strategies and challenges. Curr Pharm Des 15(6):587–600

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662

Munoz-Torrero D, Camps P (2006) Dimeric and hybrid anti-Alzheimer drug candidates. Curr Med Chem 13(4):399–422

Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ, Holmberg SD (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N Engl J Med 338(13):853–860

Pani A, Loi AG, Mura M, Marceddu T, La Colla P, Marongiu ME (2002) Targeting HIV: old and new players. Curr Drug Targets Infect Disord 2(1):17–32

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Ryckmans T, Balancon L, Berton O, Genicot C, Lamberty Y, Lallemand B, Pasau P, Pirlot N, Quere L, Talaga P (2002) First dual NK1 antagonists-serotonin reuptake inhibitors: synthesis and SAR of a new class of potential antidepressants. Bioorg Med Chem Lett 12(2):261–264

Schinazi RF, Hernandez-Santiago BI, Hurwitz SJ (2006) Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses. 71:322–334. Erratum in Antiviral Res 72 (3):256–256

Thomsen R, Christensen M (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 1:3315–3321

Wallace AC, Laskowski RA, Thornton JM (1995) Ligplot—a program to generate schematic diagrams of protein ligand interactions. Protein Eng 8(2):127–134

Yamamoto M, Ikeda S, Kondo H, Inoue S (2002) Design and synthesis of dual inhibitors for matrix metalloproteinase and cathepsin. Bioorg Med Chem Lett 12(3):375–378

Yeni P (2006) Update on HAART in HIV. J Hepatol 44:S100–S103

Young B, Kuritzkes DR (2002) Resistance to HIV-1 protease inhibitors. Infect Dis Ther 25:257–282

Computer-Aided Perspective for the Design of Flexible HIV Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs): *de-novo* Drug Design, Virtual Screening and Molecular Dynamics Simulations

Suri Moonsamy and Mahmoud E. S. Soliman*

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

Abstract: Flexible hydroquinone-based compounds were previously proposed as potential mutant-resistant NNRTIs inhibitors, however, experimental or computational evidences did not support this proposal. Herewith, using an integrated *in-silico* computational approach involving *de-novo* drug design, structure-based virtual screening (SBVS), molecular dynamics simulations and post-dynamic per-residue binding energy decomposition analysis, the binding affinity as well as the interaction landscape of novel flexible hydroquinone-based compounds were investigated to explore their activity as potential NNRTIs. The proposed leads were found to exhibit improved binding affinity when compared to FDA-approved NNRTIs, rilpvirine, nevirapine and efavirenz, however, the bioavailability profile of these compounds could hamper their uses as effective drugs. Results obtained from this extensive study could assist medicinal and biochemistry researchers with further experimental investigations.

Keywords: Flexible HIV-1 inhibitors, Non-nucleoside reverse transcriptase inhibitors, *de-novo* drug design, Virtual screening, Molecular dynamic simulations.

1. INTRODUCTION

As the human immunodeficiency virus type 1 (HIV-1) has been proven to be the cause of the acquired immunodeficiency syndrome (AIDS) [1], this disease poses detrimental effects for world health as it is estimated that 34 million people live with HIV/AIDS worldwide [2].

During the past two decades, the following two strategies have been often adopted to find drugs against AIDS. One is to target the HIV reverse transcriptase (RT) [3-6]; the other is to design HIV protease inhibitors [7-9]. The reverse transcription (RT) enzyme is important for HIV-1 replication [10]. HIV-1 RT is a vital component in the virus life cycle, where this enzyme facilitates converting single-stranded RNA viral genome into double-stranded DNA [10-12]. As RT is considered as a critical component in the virus life cycle, this HIV protein has become a major target for antiviral therapy [13, 14]. In particular, the two classes of antiviral drugs that target RT are nucleoside RT inhibitors (NRIs) and non-nucleoside RT inhibitors (NNRTIs) [10, 15, 16]. Both classes inhibit the RT-associated polymerase activity: the NRTIs compete with the natural dNTP substrate and act as chain terminators; whereas the NNRTIs bind to an allosteric pocket and inhibit polymerization non-competitively [10]. Although, a large number of inhibitors that target RT have been approved for treatment [15, 16], there are many complications that hinder the therapeutic effect of these inhibitors. For instance, NRTIs demonstrated high levels of cellular toxicity [17], whereas the highly potent and less toxic NNRTIs are severely weakened by the rapid emergence of

drug-resistant strains [1, 18]. Therefore, tackling the drug-resistant implications associated with NNRTIs could result in less toxic and more potent inhibitors as anti-HIV RT. It is strongly believed that more flexible NNRTIs could be more effective against drug-resistant mutants [11, 19, 20].

The concept of designing flexible RT inhibitors has been proposed before [11, 19-22]. Ohtaka and his colleagues made a conclusion that a protease inhibitor required to be flexible in nature in order to bind to mutant enzymes [19]. Soon afterwards, Das *et al.* (2005), concluded that conformational flexibility of an NNRTI was crucial in preserving the functionality of an NNRTI being exposed to mutations of resist antiviral strains [20]. Furthermore, it was suggested that the L100 mutant of HIV-1 RT was proficiently inhibited by diarylpyrimidine (DAPY)-based inhibitors [20], as these inhibitors are torsionally flexible and so are capable of binding to any conformation of the hydrophobic pocket of the wild-type RT and other drug-resistant mutants [23].

To this end, in this report, via using integrated *in-silico* approaches, including *de-novo* drug design, structure-based virtual screening (SBVS) and molecular dynamics simulations, we propose novel flexible hydroquinone-based structures as potential leads to inhibit HIV-1 RT. Specifically, our main purpose is to design and identify flexible ligands with enhanced conformational freedom to allow positional adaptability within RT enzyme active site and as a result could become mutant-resistant inhibitors.

Different from the approach for finding the peptide-like drugs against HIV and AIDS (see, e.g [22, 24]) based on the Chou's distorted key theory as elaborated in a comprehensive review [8], the current design/search for novel compounds was based on non-peptide structures. Although the peptide-drugs have the advantage of less toxicity [25], the

*Address correspondence to this author at the School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa; Tel: +27 031 260 7413; Fax: +27 031 260 779; E-mail: soliman@ukzn.ac.za
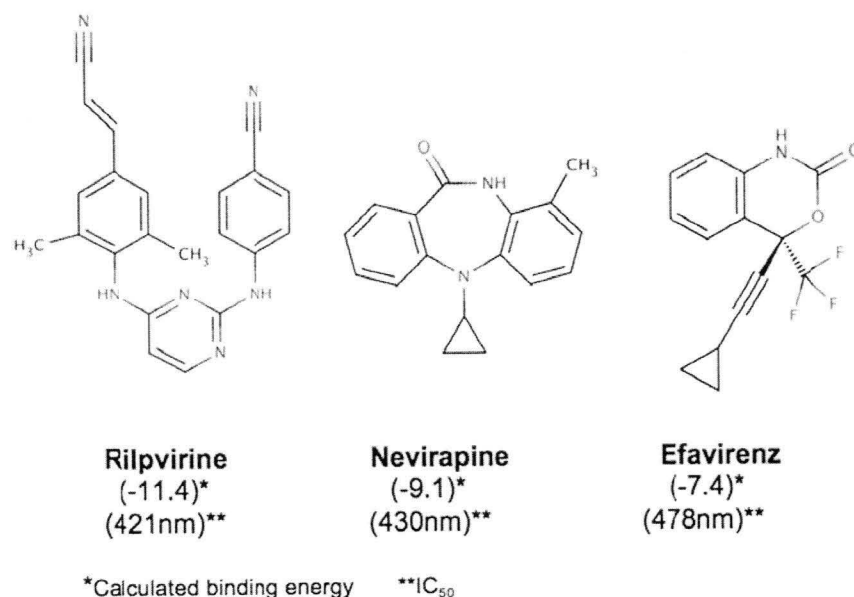
Rilpivirine
(-11.4)*
(421nm)**

Nevirapine
(-9.1)*
(430nm)**

Efavirenz
(-7.4)*
(478nm)**

*Calculated binding energy **IC_{50}

Fig. (1). 2D structures of FDA-approved HIV NNRTIs used as reference compounds for validation of docking protocol.

chemical compound drugs have the advantage of less molecular weight and easier to be absorbed [26].

The proposal of hydroquinone-based structures as flexible NNRTIs was first introduced by Bruccoleri [11], however, this proposal was not supported by either experimental or computational evidences. Therefore, the work reported herein is an attempt to verify the activity of hydroquinone-based compounds as NNRTIs, as proposed by Bruccoleri [11]. In order to accomplish this, we have designed and investigated the binding affinity and themes for a novel set of different hydroquinone-based structures against RT enzyme at molecular level. This *de-novo* drug design was then followed by structure-based virtual screening in order to identify a wider range of novel hydroquinone-based compounds from the commercially available chemical databases. Docking calculations, molecular dynamic stimulations and post-dynamics analysis and per-residue energy contribution were performed in order to accurately estimate the relative binding affinities of the proposed compounds against RT enzyme as well as to ensure the stability of the resulted ligand-enzyme complexes. To validate our calculations, the same docking approach adopted for the proposed compounds was performed for different FDA-approved NNRTIs and results were compared against experimental data.

Information gained from this study could shed light on the activity of a new series of lead compounds as potential anti-HIV RT. Also, the study could be of much interest to medicinal chemistry and biochemistry researchers for further investigation and development.

2. COMPUTATIONAL METHODS

2.1. *De-Novo* Drug Design of the Proposed Structures and Conformational Ensembles

Seven proposed structures were modeled, geometry optimized and energy minimized using MMFF94 force field implemented in Avogadro software [27]. Via conformational sampling, low-energy conformations were produced for each proposed ligand using Omega v1.8b [28]. Conformational isomers developed for each molecule were only examined, if these isomers were within a range of 3.0 kcal/mol from the lowest energy conformer.

2.2. Protein Systems

The coordinates of the RT enzyme bound to rilpvirine, a new generation NNRTI and commonly abbreviated as TMC 278, were acquired from the protein data bank (http://www.rcsb.org/) (PDB code: 2ZD1). Other NNRTIs used in this study were also obtained from the protein data bank and identified by the following PDB codes: nevirapine (NVP) (PDB code: 4B3Q) and efavirenz (EFZ) (PDB code: 1IKW) (Fig. 1).

The full computational scenario describing the preparation of enzyme systems for subsequent docking and MD simulations can be found in our earlier reports [29-33].

2.3. Docking Calculations: Structure-Based Virtual Screening (SBVS)

Docking calculations were conducted using Autodock Vina [34] screening software. The Screening procedure was run using the software default settings. The grid box used to define the screening site was elucidated using the Autodock Vina functionality built into Chimera [35]. The box was defined around Asp55, Glu56, Asp79, Arg82, Ile84, Pro85, Val99, Lys108, Val123, Arg141 and Ser169. The X, Y and Z centers were defined as 49.7, -27.3 and 38.11, respectively. The X, Y and Z size dimensions were defined as 42, 38 and 48, respectively. Autodock Vina generated screening results in the pdbqt format. Ten compounds with the best binding energies were selected from each of the two libraries. The docking results of these compounds were visualized the Viewdock feature in Chimera software.

To validate the accuracy of docking method in this work, three FDA-approved drugs with known experimental data (IC$_{50}$) were seeded in the docking runs and results were compared against experimental data.

2.4. 2D Shape Similarity-Based Compounds Library Generation

The best-docked structure resulted from the *de-novo* drug design (section 2.1) was used as template for the generation of a 2D shape similarity-based compounds library retrieved from the Chemicalize database (http://www.chemicalize.org/). The structure of the template compound was drawn using the MarvinSketch Software (Chemaxon, http://www-chemaxon.com). This was then used to generate a SMILES code, which was used to query the Chemicalize database. The database was queried for all structures that had greater than 70% shape similarity to the reference template. This results in a total number of 1007 compounds. As explained in our previous report, physicochemical filters were imposed to enhance structures mining process [36]. However, in this study, only compounds with molecular weight between 300 and 550 mg/mol will be selected – this resulted in 303 hits. No other criteria were imposed to ensure the inclusion of the maximum number of compounds. These compounds were downloaded as a single file in the mol2 format. Molegro Molecular Viewer (MMV) software suite [37] was then used to create the individual mol2 files. These files were then converted into the pdbqt format required for Autodock Vina using Raccoon software [38]. Conformational sampling (see section 2.1) was also performed for all retrieved compounds before docking into the RT enzyme.

2.5. Molecular Dynamics (MD) and Post-Dynamic Analysis

Many marvellous biological functions in proteins and DNA and their profound dynamic mechanisms, such as switch between active and inactive states [39], cooperative effects [40], allosteric transition [41], intercalation of drugs into DNA [42], and assembly of microtubules [43], can be revealed by studying their internal motions as summarized in a comprehensive review [44], and a recent paper about theoretical and experimental biology [45]. Likewise, to really understand the action mechanism of receptor-ligand binding, we should consider not only the static structures concerned but also the dynamical information obtained by simulating their internal motions or dynamic process. To realize this, the MD simulation is one of the feasible tools.

The two best docked ligand-enzyme complexes resulted from the SBVS for the *de-novo* drug designed structures and 2D shape similarity-based compounds library were subjected to MD simulations using Amber software [46] following the procedure explained in earlier reports [47-50]. For analyzing the interactions of post-dynamic protein-ligand complexes, the Ligplot programme (http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/) was used to plot the hydrogen bond and hydrophobic interactions between receptor and ligand atoms within a range of 5 Å. Chimera software [35] was used to visualize and prepare figures for the top ranked molecules (Fig. **5A** & **5B**). Per-residue energy decomposition was computed using Moldock scoring functions [51].

3. RESULTS AND DISCUSSION

3.1. *De-novo* Drug Design Plan

First, the enzyme target, RT enzyme, active site was carefully examined in order to explore the nature of the binding site pocket. It was observed that the binding themes differ amongst the crystal structures of various NNRTIs to RT. Specifically, NNRTIs bind to the hydrophobic pocket situated in the p66 subunit of HIV-1 RT [11]. In our study, we determined the active site residues of RT to include: Lys-103, Leu-236, Val-108 and Tyr-190. The NNRTI-binding site also has a high degree of flexibility and as a result can bind NNRTIS of diverse shapes and sizes, which only opens in the presence of an inhibitor [11, 20]. Some NNRTIs display "butterfly-like" conformations whereas others show "horseshoe" conformations [11, 20].

Based on the study reported by Bruccoleri [11] we proposed seven *de-novo* drug ligands (Table **1**) that we anticipate to have potent flexible non-nucleoside inhibitory activity against HIV-1 RT.

Our design initially began with a covalently linked hydroquinone trimer, structure **1** (Table **1**). Then, in order to provide more flexibility to the structure, we introduced O atom and CH$_2$ spacers between the hydroquinone rings. After docking, we noticed that these compounds do not optimally fill the enzyme active site pocket as evident from the poor binding affinity (structures **1**, **2** and **3** - Table **1**). Then we opted to design larger structures by adding phenyl rings on the spacer atoms, CH$_2$, in order to maximize the hydrophobic interactions between the inhibitors and the active site residues (structure **4**, Table **1**). As expected, this approach has improved the binding affinity by 1.2 kcal/mol (Table **1**). We tried a parallel approach to the former one, however, this time by linearly expanding the hydroquinone moieties to create tetramers with and without spacers, compounds **5**, **6**, respectively (Table **1**). However, these compounds showed modest activity when compared to the prototype, rilpivirine. Adding one more hydroquinone moiety to structure **7** has remarkably improved the binding affinity (Table **1**).

The reason for this observation is perhaps due to the fact that compound **7** is fully occupying the active site, therefore showed more enhanced binding affinity towards RT.

As evident from Table **1**, all the *de-novo* drug set 1 designed compounds (Table **1**), following the design plan proposed by Bruccoleri [11], exhibited lower binding affinity towards RT when compared to rilpivirine (-11.4 kcal/mol) – excluding compound **7** which shows a slightly better binding affinity, however, this is very subtle difference if experimental errors are considered.

From these results, it seems to us that the designed compounds exhibit humble activity as NNRTIs. Many factors might contribute to these findings: (1) the docking method applied might not be accurate enough to provide accurate estimates of the binding energies, (2) our chemical intuition might have drifted away and led to the design in the wrong direction (3) our design have not covered a wider range of structural scaffolds which might have higher binding affinities towards RT. In order to eliminate the first factor, we

Table 1. The 2D structure, calculated binding energy and physicochemical properties of the *de-novo* drug designed compounds as well as the top-ten best-docked compounds retrieved from chemical database.
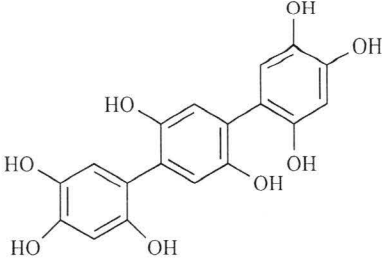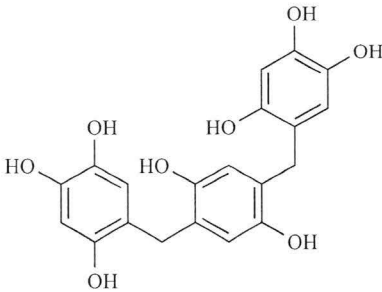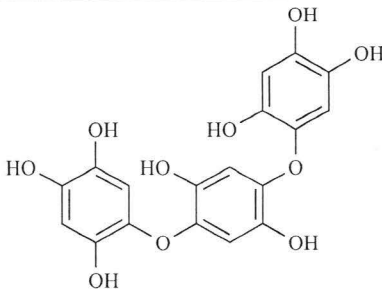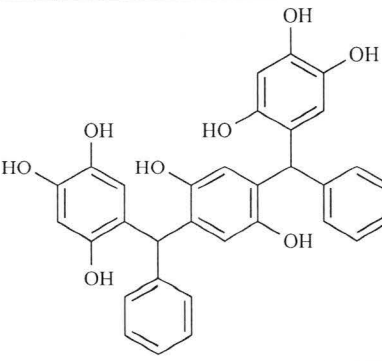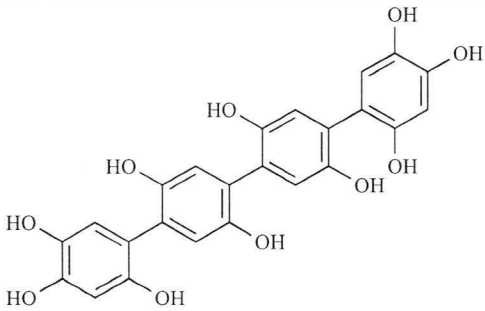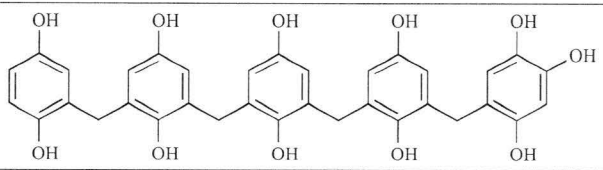
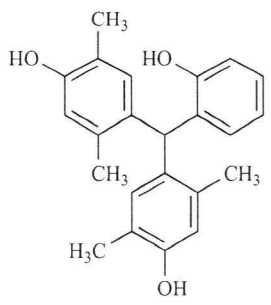| Compound	2D Structure	Docked Binding Energy (kcal/mol)	logP	Molecular Weight (g/mol)
De-novo drug set 1				
1		-8.2	2.8	358.3
2		-8.8	3.7	386.1
3		-7.7	2.5	390.3
4		-10.0	3.7	544.6
5		-7.5	3.9	466.3

Table 1. contd...
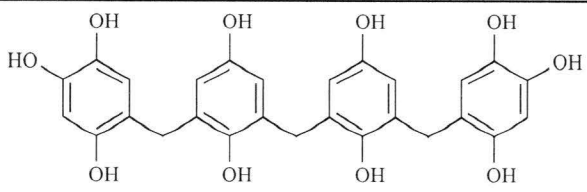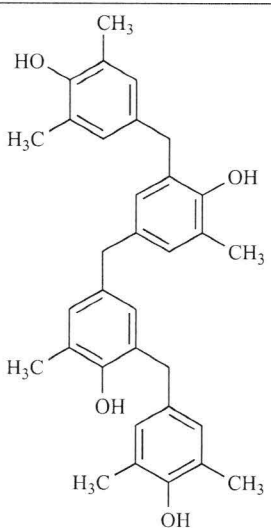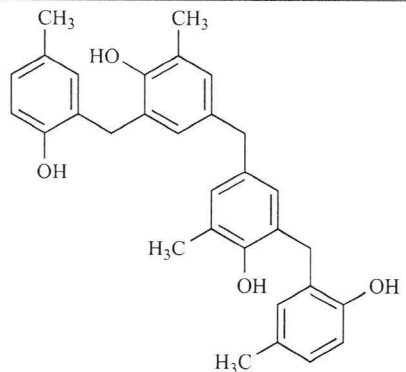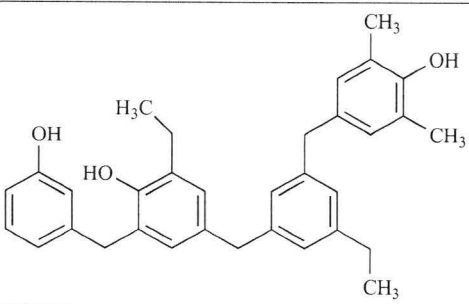
Compound	2D Structure	Docked Binding Energy (kcal/mol)	logP	Molecular Weight (g/mol)
6		-10.5	4.4	538.5
7		-11.9	7.3	598.6
Compounds library from chemical database				
8		-12.8	10.11	496.64
9		-12.8	9.09	468.55
10		-12.4	11.00	524.69

Table 1. contd…
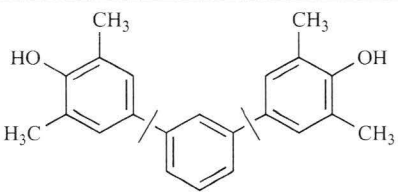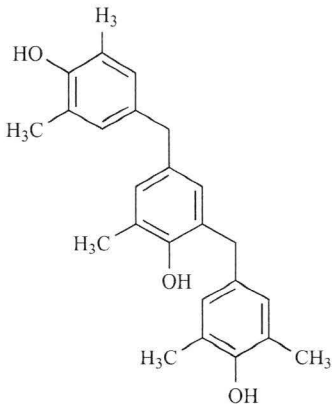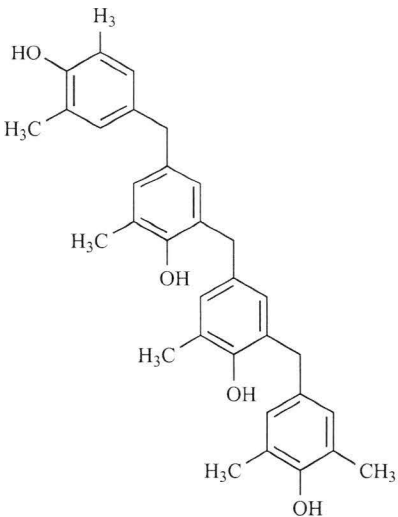
Compound	2D Structure	Docked Binding Energy (kcal/mol)	logP	Molecular Weight (g/mol)
11		-12.3	6.63	348.43
12		-12.2	8.78	402.57
13		-12.0	7.81	376.49
14		-11.9	10.11	496.64
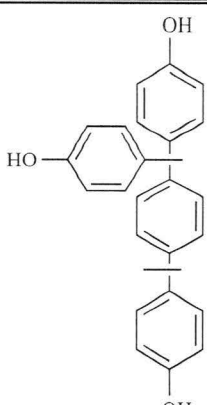
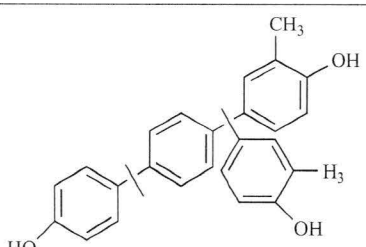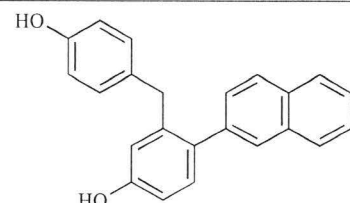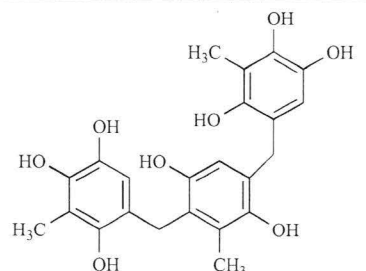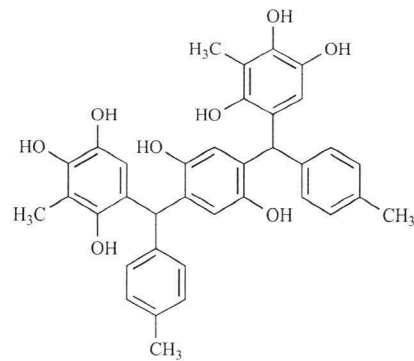| Compound	2D Structure	Docked Binding Energy (kcal/mol)	logP	Molecular Weight (g/mol)
15		-11.8	7.55	424.53
16		-11.8	8.58	452.58
17		-11.8	6.17	340.41
De-novo drug set 2				
18		-9.7	5.27	428.43
19		-10.9	8.62	594.23

Table 1. contd…

Compound	2D Structure	Docked Binding Energy (kcal/mol)	logP	Molecular Weight (g/mol)
20		-11.1	7.27	564.58

opted to validate the docking protocol applied in this work. One of the most commonly used approaches to validate docking results is "cross-validation" where other docking programs are used to validate the data obtained from the original docking software. From our experience, we found that using different software, with different scoring functions, could generate different results and be misleading, therefore, we did not opt to use this validation method. We always believe that the most rational way to validate docking calculations, or even any other computational tool, is to run the calculations on a set of compounds with available experimental data. Results from calculations, are then compared against experimental data for validation. Another reliable approach to validate docking results is to perform docking for the bound inhibitors in the X-ray crystal structure of the inhibitor-enzyme complex, then the RMSD of the docked complex is then compared against the native X-ray structure.

To this end, in this work, we performed docking analyses on three FDA approved NNRTI drugs using our docking methodology (Fig. 1). The PDB coded structures were obtained from the Protein Data Bank (http://www.rcsb.org/) and experimental IC_{50} values were obtained from the Binding database (http:// www.bindingdb.org/). It is worth mentioning that, we have previously used the same docking approach in several experimental reports and it proved reliable to a great extent [29-33]. As evident from the docking results (Fig. 1), the docked energies are in great accordance with the experimental IC_{50}. This implies that the docking methods used in this work could be reliable enough to estimate the binding affinities for the proposed compounds.

In light of the consistency between the calculated binding energies and experimental IC_{50} values, our docking approach proved to be reproducible and valid. Furthermore, the root mean squared deviation (RMSD) between the docked structure of Nevirapine against the X-ray crystal structure of Nevirapine with RT (PDB code: 3V81) is 0.79 Å, proving that the docking approach is also reliable enough to predict the ligand orientation in the enzyme active site.

As mentioned above, and as evident from the humble binding affinities of the *de-novo* drug designed compounds, our chemical intuition for the design of potential leads might have failed to build up the right structural scaffolds or at least have not covered a wider range of compounds to probably include those with better binding affinities. To this end, we opted to search for similar compounds in the commercially available chemical database (www.chemicalize.org). Our query of the database was based on 2D

shape similarity-based identity to our best-docked structure (compound 7- Table 1) from the *de-novo* design as a reference template. All compounds in the generated library (see Methods sections for details) were then docked into the active site of the RT enzyme. As shown in Table 1, the top 10 compounds from the SVBS session exhibited remarkably higher binding energies as compared to Rilpvirine (-11.4 kcal/mol), with binding energies ranging from -12.8 to -11.8 kcal/mol.

By looking at the structures retrieved from the database, we found that, however, architecturally similar, none of the compounds retained a hydroquinone fragment. Also, in all structures, a carbon spacer exists between all phenyl rings. Another conclusion can be drawn by looking at Table 1, the *de-novo* drug designed library *vs.* virtually screened library, is that inclusion of methyl groups at different positions on the main structure skeleton has a significant positive impact on the binding affinity. This could be explained based on the fact that the NNRT active site is mostly hydrophobic in nature. Based on the conclusions drawn from Table 1, we opted to optimize the initially *de-novo* drug designed leads by inclusion of methyl groups on the skeleton of the structures having a spacer between aromatic rings, structure **2**, **4** and **6**, in order to see whether or not this will enhance the binding affinities. Inclusion of methyl groups on the hydroquinone skeleton, structure **18**, **19** and **20**, has improved the binding affinity by 0.5 - 0.9 kcal/mol (Table 1). We assume that inclusion of methyl group has increased the chances of hydrophobic interactions with the hydrophobic residues in the enzyme active site. To verify this assumption, we performed interaction and per-residue binding energy decomposition analysis on the post-dynamic ligand enzyme complexes resulted from the docking calculations (see section 3.2).

3.2. Molecular Dynamics Simulations and Post-Dynamic Analysis

In order to obtain more insight on the stability of the resulted docked complexes, the nature of overall interaction theme between the proposed ligands and the target protein and the specific amino acids involved in the ligand binding, we performed a 1 ns MD simulations followed by extensive post-dynamic analyses on the docked ligand-enzyme complexes from the *de-novo* drug design and chemical database.

From our previous experience with molecular docking, in many occasions, we experienced that even best docked structures may fly away from the enzyme active site within a few picoseconds of MD simulations. Therefore, we believe that
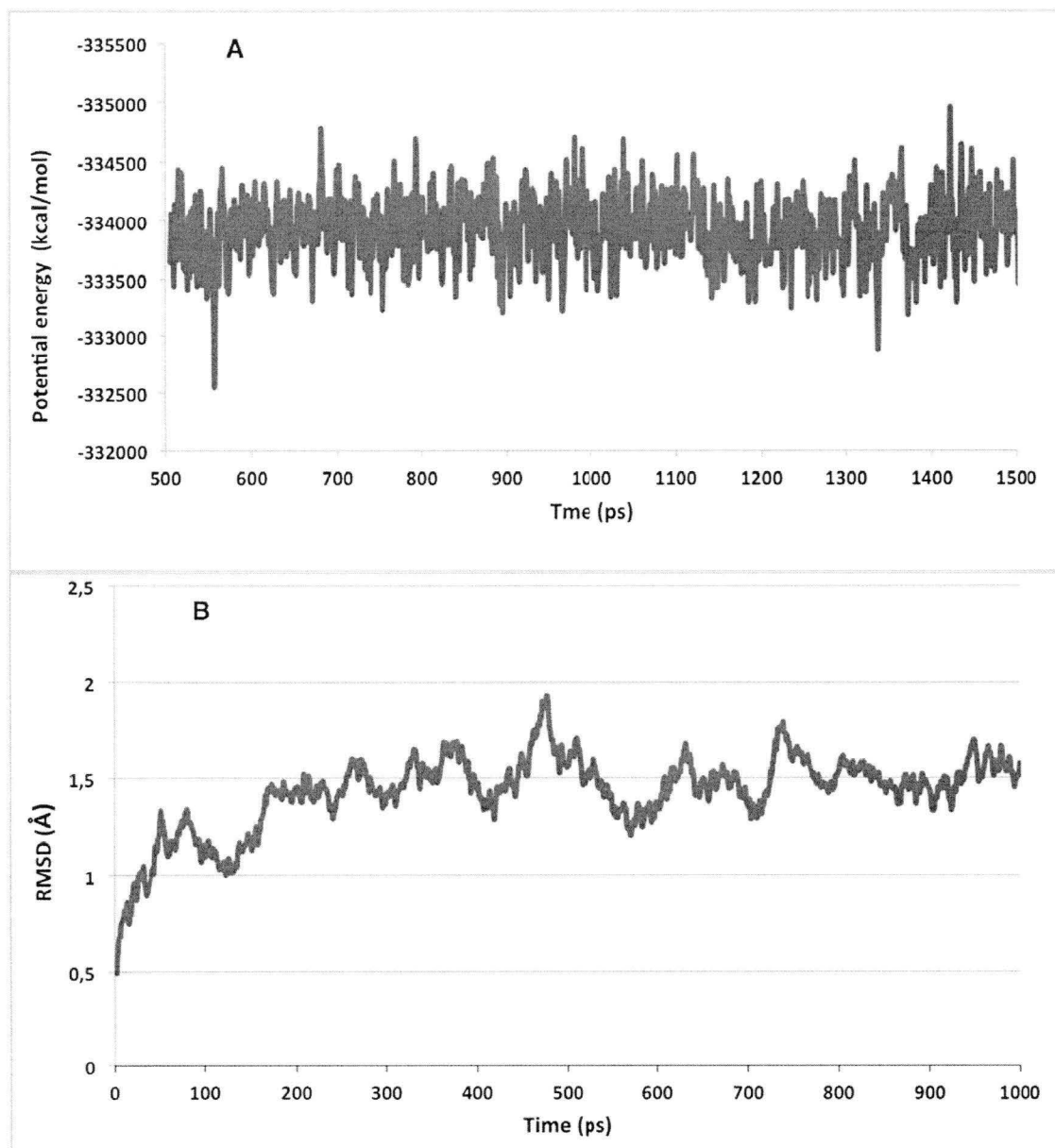
Fig. (2). Potential energy and RMSD fluctuation over the last 1 ns of the MD simulations.

docking calculations that are not validated by relatively a long MD run to ensure the stability of the system might not be reliable. Interestingly, for all the compounds-enzyme complexes, the average RMSD values were below 1.5 Å and the variation of the potential energies falls within 900 kcal/mol and this is a good indication of the system stability (Fig. **2**). The PDBs for all docked complexes are provided with the supplementary material.

In order to perform additional lead optimization and expansion phases, it is highly beneficial to have a quantitative estimation of the contribution of individual amino acid residues in the enzyme active site towards ligand binding and we achieved this by the computation of the per-residue interaction MolDock scoring function [51]. Per-residue interactions for the reference compound (TMC 278) and compounds showed improved affinities with RT, compounds **8**, **7**, and **20** are shown in Fig. (**3**).

From the per-residue interaction calculations, compounds **7**, **8** and **20** demonstrated remarkable interactions with the active site residues Tyr183, Pro97, Leu102, Leu236, Phe229 and Tyr190 (Fig. **2**). However, to further gain better depiction of these interactions, hydrogen bond and electrostatic interactions maps for the 2n MD averaged complex structures of **8** and **20** with RT (Fig. **4**). These plots were created using Ligplot software (http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/).

As evident from the post-dynamic hydrogen bond and electrostatic interactions (Fig. **4A**) as well as per-residue analysis (Fig. **3**), the *de-novo* drug design-based compound 20 showed a strong hydrophobic interaction with Tyr183 and hydrogen bond interactions with His237, Lys103, Gly95 and Gln93 (Fig. **4A**). Table **1** and Fig. (**4**) also showed that compound 8 exhibited its favorable affinity than compound 20
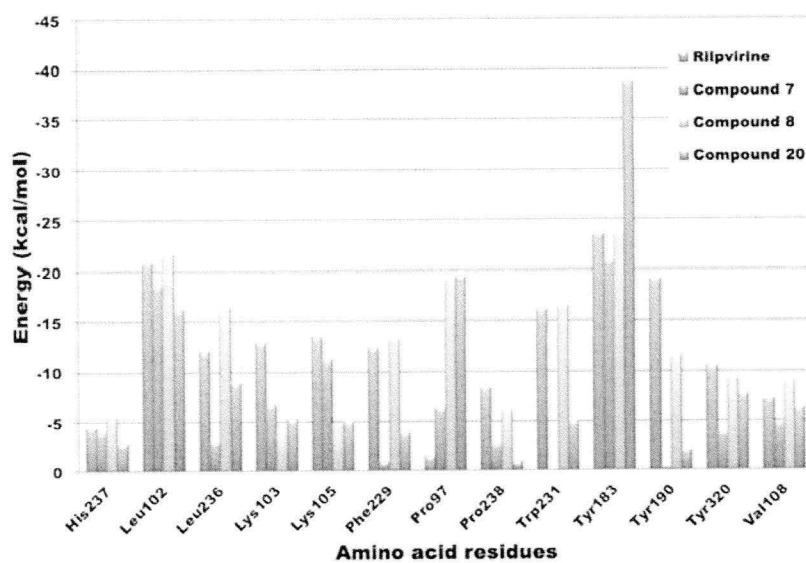
Fig. (3). Per-residue interactions of the reference compound (TMC 278) and the proposed compounds 7, 8, and 20 with RT.
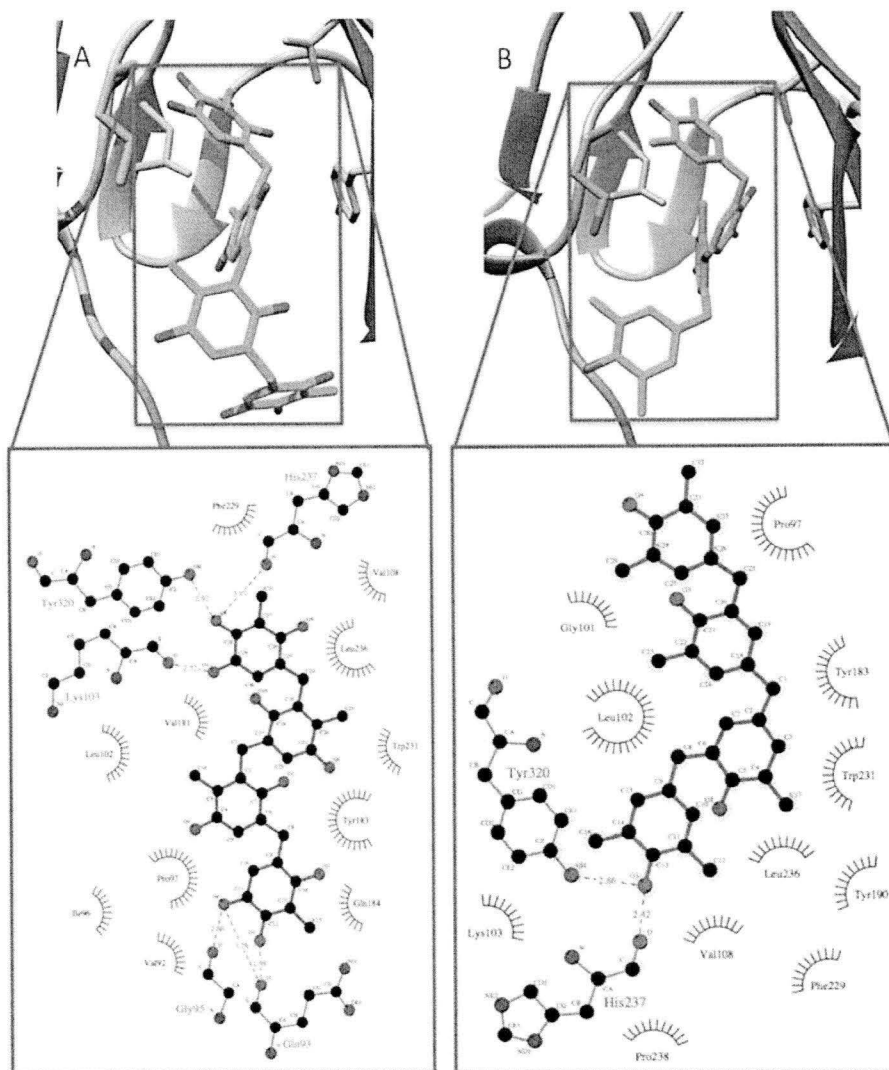

Fig. (4). Hydrogen bonding and electrostatic interactions for the *de-novo* drug design-based compound 20 in complex with RT (A) and compound 8 with RT (B), respectively.

due to the increased hydrophobic interactions with the RT active site residues.

The integrated computational tools adopted in this work has explored the inhibitory landscape of a number of hydroquinone-based polymeric compounds as suggested by Bruccoleri [11] and proved that these compounds, to a reasonable extent, could be potential leads as anti HIV RT. However, the bioavailability profile of these compounds could be problematic since most of them violates Lipinski's rule of five (especially, logP values and the number of hydrogen bond donors and acceptors). Also the toxicity profiles of these compounds need to be experimentally evaluated.

We believe that the computational workflow presented in this study could be a useful tool in the process of design and development of novel leads against biological targets

4. CONCLUSION

In the present study, and guided by the study presented by Bruccoleri [11], we proposed seven *de-novo* hydroquinone -based ligands that we anticipated to have potent flexible non-nucleoside inhibitory activity against HIV-1 RT. However, these *de-novo* proposed compounds exhibited moderate binding affinities when compared to a rilpivirine. To optimize our *de-novo* drug design, we opted to mine for a wider range of similar compounds from the chemical database. However, the chemical database did not retrieve any hydroquinone-based structures, it guided us to optimize our leads. Inclusion of hydrophobic moieties on the hydroquinone skeleton has improved the binding interaction with the RT active site. Despite that fact that proposed compounds exhibited comparable binding affinities as rilpivirine, most of these compounds violate Lipinski's rule of five and expected to have low bioavailability profile. Along with the computational insight provided from this report, further experimental investigations are needed to evaluate the bioavailability as well as the toxicity profiles of these compounds.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

SUPPLEMENTARY MATERIAL

The PDB coordinates for all docked compounds RT complexes are provided as supplementary material.

REFERENCES

[1]　Wainberg MA, S.J.; Montaner, J.S.; Murphy, R.L.; Kuritzkes, D.R.; and Raffi, F. Challenges for the clinical development of new nucleoside reverse transcriptase inhibitors for HIV infection, *Antivir. Ther.*, **2005**, *10*, 13-28.

[2]　Johnson, B.C.; Pauly, G.T.; Rai, G.; Patel, D.; Bauman, J.D.; Baker, H.L.; Das, K.; Schneider, J.P.; Maloney, D.J.; Arnold, E.; Thomas, C.J.; and Hughes, S.H. A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants. *Retrovirology*, **2012**, *9*, 1-11.

[3]　Althaus, I.W.; Chou, J.J.; Gonzales, A.J.; Deibel, M.R.; Chou, K.C.; Kezdy, F.J.; Romero, D.L.; Palmer, J.R.; Thomas, R.C.; Aristoff, P.A.; Tarpley, W. G.Reusser, F. Kinetic-studies with the nonnucleoside HIV-1 reverse-transcriptase inhibitor-U-88204E. *Biochemistry*, **1993**, *32*, 6548-6554.

[4]　Chou KC, K.F., Reusser F Review: Steady-state inhibition kinetics of processive nucleic acid polymerases and nucleases. *Anal. Biochem.*, **1994**, *221*, 217-230.

[5]　Althaus, I.W.; Gonzales, A.J.; Chou, J.J.; Romero, D.L.; Deibel, M. R.; Chou, K.C.; Kezdy, F.J.; Resnick, L.; Busso, M.E.; So, A.G.; Downey, K.M.; Thomas, R.C.; Aristoff, P.A.; Tarpley, W.G.; Reusser, F. The quinoline U-78036 is a potent inhibitor of HIV-1 reverse-transcriptase. *J. Biol. Chem.*, **1993**, *268*, 14875-14880.

[6]　Althaus, I.W; Chou, J.J.; Gonzales, A.J.; Diebel, M.R.; Chou, K.C.; Kezdy, F.J.; Romero, D.L.; Aristoff, P.A.; Tarpley, W.G.; Reusser, F. Steady-state kinetic studies with the non-nucleoside HIV-1 reverse transcriptase inhibitor U-87201E. *J. Biol. Chem.*, **1993**, *268*, 6119-6124.

[7]　Chou, K.C.A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. *J. Biol. Chem.*, **1993**, *268*, 16938-16948.

[8]　Chou, K.C. Prediction of human immunodeficiency virus protease cleavage sites in proteins. *Anal. Biochem.*, **1996**, *233*, 1-14.

[9]　Shen, H.-B.Chou, K.-C. HIVcleave: a web-server for predicting human immunodeficiency virus protease cleavage sites in proteins. *Anal. Biochem.*, **2008**, *375*, 388-390.

[10]　Esposito, F., Corona, A., and Tramontano, E. HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Mol. Biol. Int.*, **2012**, *2012*, 1-23.

[11]　Bruccoleri, A. Positional Adaptability in the Design of Mutation-Resistant Nonnucleoside HIV-1 Reverse Transcriptase Inhibitors: A Supramolecular Perspective. *AIDS Res. Hum. Retrovi.*, **2013**, *29*, 1-9.

[12]　Seckler, J.M., Howard, K.J., Barkley, M.D., and Wintrode, P.L. Solution Structural Dynamics of HIV-1 Reverse Transcriptase Heterodimer. *Biochemistry*, **2009**, *48*, 7646-7655.

[13]　Ilina, T., LaBarge, K., Sarafianos, S.G., Ishima, R., and Parniak, M.A. Inhibitors of HIV-1 Reverse TranscriptaseAssociated Ribonuclease H Activity. *Biology*, **2012**, *1*, 521-541.

[14]　Harpstead, B. HIV-1 reverse transcriptase structure, activities, and inhibitors. *Basic Biotechnol.*, **2005**, *MMG 445*, 1-7.

[15]　Bauman, J. D., Das, K., Hol, W.C., Baweja, M., Himmel, D.M., Clark Jr,A.D., Oren, D.A., Boyer, P.L., Hughes, S.H., Shatkin, A,J., and Arnold, E. Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acids Res.*, **2008**, *36*, 5083-5092.

[16]　Pani, A., Loi, A.G., Mura, M., Marceddu, T., La Colla, P., and Marongiu, M. E. Targeting HIV: Old and new players. *Curr. Drug Tar. Infe. Disord.*, **2002**, *2*, 17-32.

[17]　Schinazi, R.F., Hernandez-Santiago, B.I., and Hurwitz, S.J. Pharmacology of current and promising nucleosides for the treatment of human immunodeficiency viruses. *Antivir. Res.*, **2006**, *71*, 256.

[18]　Gallant, J.E.; Gerondelis, P.Z.; Wainberg, M.A.; Shulman, N.S.; Haubrich, R.H.; St Clair, M.; Lanier, E.R.; Hellmann, N.S.; and Richman, D.D. Nucleoside and nucleotide analogue reverse transcriptase inhibitors: A clinical review of antiretroviral resistance. *Antiviral Ther.*, **2003**, *8*, 489-506.

[19]　Ohtaka, H.; V.-C. A.; Xie, D.; and Freire E. Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease. *Protein Sci.*, **2002**, *11*, 1908-1916.

[20]　Das K, L.P.; Hughes, S.H.; and Arnold, E. Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Prog Biophys Mol. Biol.*, **2005**, *88*, 209-231.

[21] Du, Q.-S.; Sun, H.; Chou, K.-C. Inhibitor design for SARS coronavirus main protease based on "Distorted key theory". *Med. Chem.,* **2007,** *3,* 1-6.

[22] Gan, Y.R.; Huang, H.; Huang, Y.D.; Rao, C.M.; Zhao, Y.; Liu, J.S.; Wu, L.Wei, D.Q. Synthesis and activity of an octapeptide inhibitor designed for SARS coronavirus main proteinase. *Peptides,* **2006,** *27,* 622-625.

[23] Sarafianos S, *e. a.* Taking aim at a moving target: Designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases. *Curr. Opin Struct. Biol.,* **2004,** *14,* 716-730.

[24] Du, Q. S.; Sun, H.Chou, K.C. Inhibitor design for SARS coronavirus main protease based on "Distorted key theory". *Med. Chem.,* **2007,** *3,* 1-6.

[25] Chou, K.C. Structural bioinformatics and its impact to biomedical science. *Curr. Med. Chem.,* **2004,** *11,* 2105-2134.

[26] Du, Q. S.Huang, R.B. Recent Progress in Computational Approaches to Studying the M2 Proton Channel and Its Implication to Drug Design Against Influenza Viruses. *Curr. Protein Pept. Sci.,* **2012,** *13,* 205-210.

[27] Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform". *J. Cheminform.,* **2012,** *4,* 1-17.

[28] Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B. A.Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inform. Model.,* **2010,** *50,* 572-584.

[29] Makatini, M.M.; Petzold, K.; Sriharsha, S.N.; Ndlovu, N.; Soliman, M.E.; Honarparvar, B.; Parboosing, R.; Naidoo, A.; Arvidsson, P. I.; Sayed, Y.; Govender, P.; Maguire, G.E.; Kruger, H. G.; Govender, T. Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. *Eur. J. Med. Chem.,* **2011,** *46,* 3976-3985.

[30] Makatini, M.M.; Petzold, K.; Sriharsha, S.N.; Soliman, M.E.; Honarparvar, B.; Arvidsson, P.I.; Sayed, Y.; Govender, P.; Maguire, G.E.; Kruger, H. G.Govender, T. Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease. *Bioorg. Med. Chem. Lett.,* **2011,** *21,* 2274-2277.

[31] Karpoormath, R.; Sayed, Y.; Govender, P.; Govender, T.; Kruger, H. G.; Soliman, M. E.Maguire, G. E. 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.,* **2012,** *40,* 19-29.

[32] Honarparvar, B.; Makatini, M.M.; Pawar, S.A.; Petzold, K.; Soliman, M.E.; Arvidsson, P.I.; Sayed, Y.; Govender, T.; Maguire, G.E. Kruger, H.G. Pentacycloundecane-diol-Based HIV-1 Protease Inhibitors: Biological Screening, 2D NMR, and Molecular Simulation Studies. *ChemMedChem.,* **2012,** *7*(6), 1009-1019.

[33] Makatini, M.M.; Petzold, K.; Alves, C.N.; Arvidsson, P.I.; Honarparvar, B.; Govender, P.; Govender, T.; Kruger, H.G.; Sayed, Y.; Jeronimolameira; Maguire, G.E. Soliman, M.E. Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors. *J. Enz. Inhibit. Med. Chem.,* **2013,** *28*(1), 78-88.

[34] Trott, O.Olson, A.J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.,* **2010,** *31,* 455-461.

[35] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.,* **2004,** *25,* 1605-1612.

[36] Soliman, M.E. A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors. *Drug Dev. Res.,* **2013,** *74*(5), 283-295.

[37] CLCBio. http://www.clcbio.com/products/molegro/

[38] Forli, S. AutoDock | Raccoon: an automated tool for preparing AutoDock virtual screenings, http://autodock.scripps.edu/resources/raccoon.

[39] Wang, J.-F.; Chou, K.-C. Insight into the molecular switch mechanism of human Rab5a from molecular dynamics simulations. *Biochem. Biophys. Res. Commun.,* **2009,** *390,* 608-612.

[40] Chou, K.C. Low-frequency resonance and cooperativity of hemoglobin. *Trends Biochem. Sci,* **1989,** *14,* 212-212.

[41] Chou, K.C. The biological functions of low-frequency vibrations (phonons). 6. A possible dynamic mechanism of allosteric transition in antibody molecules. *Biopolymers,* **1987,** *26,* 285-295.

[42] Chou, K. C.Mao, B. Y. Collective motion in DNA and its role in drug intercalation. *Biopolymers,* **1988,** *27,* 1795-1815.

[43] Chou, K.C.; Zhang, C.T.; Maggiora, G.M. Solitary wave dynamics as a mechanism for explaining the internal motion during microtubule growth. *Biopolymers,* **1994,** *34,* 143-153.

[44] Chou, K.C. Low-frequency collective motion in biomacromolecules and its biological functions. *Biophys. Chem.,* **1988,** *30,* 3-48.

[45] Lin SX, L.J. Theoretical and experimental biology in one —A symposium in honour of Professor Kuo-Chen Chou's 50th anniversary and Professor Richard Giegé's 40th anniversary of their scientific careers. *J. Biomed. Sci. Eng. (JBiSE).,* **2013,** *6,* 435-442.

[46] Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.,* **2005,** *26,* 1668-1688.

[47] Ahmed, S.M.; Kruger, H.G.; Govender, T.; Maguire, G.E.; Sayed, Y.; Ibrahim, M. A.; Naicker, P.Soliman, M. E. 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.,* **2013,** *81,* 208-218.

[48] Karpoormath, R.; Sayed, Y.; Govender, P.; Govender, T.; Kruger, H.G.; Soliman, M.E.; Maguire, G.E. 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.,* **2012,** *40,* 19-29.

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

[50] Naicker, P.; Achilonu, I.; Fanucchi, S.; Fernandes, M.; Ibrahim, M. A.; Dirr, H.W.; Soliman, M.E.; Sayed, Y. 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.,* **2012,** *31*:12, 1370-1380.

[51] Thomsen, R.; Christensen, M.H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.,* **2006,** *49,* 3315-3321.

molecules

Article

Integrated Computational Tools for Identification of CCR5 Antagonists as Potential HIV-1 Entry Inhibitors: Homology Modeling, Virtual Screening, Molecular Dynamics Simulations and 3D QSAR Analysis

Suri Moonsamy, Radha Charan Dash and Mahmoud E. S. Soliman *

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

* Author to whom correspondence should be addressed; E-Mail: soliman@ukzn.ac.za;
  Tel.: +27-031-260-7413; Fax: +27-031-260-779.

Abstract: Using integrated *in-silico* computational techniques, including homology modeling, structure-based and pharmacophore-based virtual screening, molecular dynamic simulations, per-residue energy decomposition analysis and atom-based 3D-QSAR analysis, we proposed ten novel compounds as potential CCR5-dependent HIV-1 entry inhibitors. Via validated docking calculations, binding free energies revealed that novel leads demonstrated better binding affinities with CCR5 compared to maraviroc, an FDA-approved HIV-1 entry inhibitor and in clinical use. Per-residue interaction energy decomposition analysis on the averaged MD structure showed that hydrophobic active residues Trp86, Tyr89 and Tyr108 contributed the most to inhibitor binding. The validated 3D-QSAR model showed a high cross-validated r_{cv}^2 value of 0.84 using three principal components and non-cross-validated r^2 value of 0.941. It was also revealed that almost all compounds in the test set and training set yielded a good predicted value. Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors and serve as a powerful tool in the drug design and development machinery.

1. Introduction

The Human Immunodeficiency Virus type 1 (HIV-1) infection, the causative agent of Acquired Immunodeficiency Syndrome (AIDS) [1], still remains a fatal human health-threatening disease [2]. An estimated 34 million people live with HIV/AIDS worldwide [3,4]. The overall global estimate is that approximately 22.9 million of these individuals live in sub-Saharan Africa [3,4].

In AIDS therapy, the fundamental strategy is to inhibit viral replication and hence, to slow down the destruction of the immune system and prolong the lives of infected individuals. Currently, a number of viral targets are being used to develop anti-HIV drugs; which are essential for viral replication and survival, and these include, protease enzyme (PR) [5], reverse transcriptase (RT) [6] and integrase (IN) [7]. Several drugs which are currently in clinical use have been developed to inhibit these potential viral targets, and such as integrase inhibitors, reverse transcriptase inhibitors and protease inhibitors [8].

Numerous concerns regarding the long-term side effects of antiretroviral drugs and the increasing transmission of resistant variants accentuates the requirement to identify new classes of drugs, which are able to efficiently suppress HIV-1 replication [2]. Therefore, there is an on-going need for novel therapeutics, which can prevent the entry of HIV-1 into its target cells [9,10].

The entry of the HIV virus into its target cell is mediated by the specific interactions of the target cell itself, such as the interaction between gp120 viral envelope glycoprotein and the plasmatic membrane receptors [11]. In turn, these specific interactions produce conformational alterations in both the glycoprotein and in the membrane receptors that facilitates fusion of the HIV virus and the target cell. Numerous studies have evaluated the role of CD4 and its interaction with gp120 and concluded that the CD4-gp120 interaction is a crucial component, but is not adequate for the disease to become established [12,13].

Besides CD4, recently certain chemokine receptors (CCRs) belonging to the G-protein coupled receptor superfamily (GPCRs) have been identified as co-targets crucial for viral entry into target cells [9,14]. Different CCRs and counterpart chemokine ligands (RANTES, MIP-1alpha and MCP2) are responsible for signaling regulation within immune cells and therefore are potential target systems for preventing virus-cell fusion. Several studies have reported on the identification and characterization of diverse CCRs [15]. Besides a single CCR that is viral strain-dependent, the majority of CCR strains are R5 isolates, which are transmitted during sexual intercourse and act on CCR5 throughout the disease [15].

CCR5 has proven to be an important pharmaceutical target in the contexts of HIV-1 and other inflammatory diseases. This chemokine receptor functions as an integral protein in HIV-1 entry into host cells by acting as a crucial co-receptor for the gp120 viral envelope glycoprotein [1]. Furthermore, experimental data revealed the importance of CCR5 in HIV-1 transmission. It was reported that individuals that are homozygous for the 32-base pair deletion for the CCR5 allele produce a defective CCR5 co-receptor and are resistant to R5-tropic HIV-1 infection, however are otherwise generally healthy [16]. This fact has been an instigating factor in the past decade for identifying anti-HIV agents that specifically targets CCR5-mediated entry mechanism. Furthermore, this implies that functional inhibition of CCR5 may help protect against infection without provoking damage to patients. Thus, blocking viral entry using small-molecule antagonists selective for CCR5 might provide a new and more effective type of anti-HIV drug.

Although, the concept of designing small molecule CCR5 antagonists has been investigated before [1,10,17]; to date, no structural information about the precise binding site of CCR5 with any FDA-approved inhibitor is available. Several studies have reported CCR5 modeling of "potential" hit leads using computational approaches, including virtual screening, molecular docking, molecular dynamic stimulations and pharmacophore modeling. Perez-Nueno *et al.*, reported a detailed comparative report of ligand-based and receptor-based virtual screening methods to unveil potential HIV entry inhibitors for CXCR4 and CCR5 receptors [18]. It has been documented that structure-based virtual screening methods yield better results as compared to ligand-based approaches. Afantitis *et al.* and Aher *et al.* identified CCR5 antagonists derived from 1-(3,3-diphenylpropyl)-piperidinyl amides using virtual screening and quantitative structure-activity relationships (QSAR) studies [19,20]. In a previous report by Kellenberger *et al.* structure-based techniques were used to model the physics of protein-ligand interactions in conjugation with combined 2D and 3D structure-based techniques [21]. Researchers have developed new approaches of combining computational molecular modeling methodologies, for example, molecular docking, 3D-QSAR, comparative receptor modeling and virtual screening to discover potential CCR5 HIV-1 entry inhibitor drugs [1]. Xu *et al.* studied the detailed interactive relationship between CCR5 and its inhibitors using a docking-based/ 3D-QSAR strategy along with protein modeling and MD stimulation [2]. However, in other mechanistic studies of protein-ligand entry inhibitor interactions, investigators have used homology modeling, molecular docking and molecular dynamic stimulation techniques [1].

To this end, in this report, via hybrid structure-based and ligand-based virtual screening, we aim to identify novel CCR5 antagonists as potential HIV-1 entry inhibitors. A human CCR5 homology model template and maraviroc, a known FDA-approved CCR5 antagonist (Figure 1), were used as prototypes. Virtual screening of ligand-based compound libraries were generated via two distinct yet complimentary approaches: (a) structural similarity-based compound generated library—this library generated compounds that bear a 2D structural similarity to the reference drug maraviroc, whereas the (b) Pharmacophore-based generated library—this library generated compounds that contained the pharmacophoric features of the reference drug structure. Merging these independent compound libraries allowed us to ensure that our generated hit lead library encompassed structural units with diversity, yet with mutual pharmacophoric and structural features. Furthermore, docking calculations were computed using the generated ligand-based compound libraries against the CCR5 enzyme.

Figure 1. 2D Structure of the known FDA-approved CCR5 antagonist maraviroc.
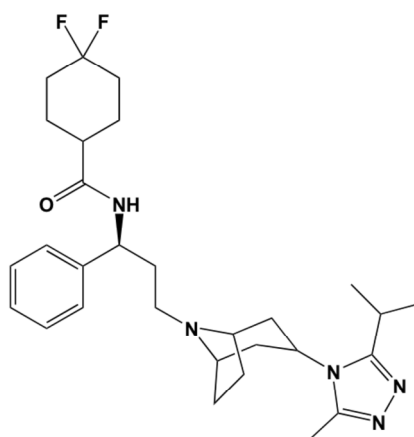
To validate our docking calculations, the same docking approach adopted for the ligand-based libraries was then performed on a set of compounds with known experimental data obtained from inhibition assays against HIV-1 CCR5 and these results were compared against experimental data. Since molecular docking may not be a true reflection of the stability of an ligand-enzyme complex, therefore, in order to gain more insight into the stability of the resulted docked systems, the nature of the overall interaction themes between the generated ligands and the target protein, and the specific amino acids involved in ligand binding, we performed 1 ns MD simulations followed by extensive post-dynamic analysis.

We took our study a step further by obtaining a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC_{50} (mM) data taken from literature for the development of our atom-based 3D-QSAR model [1].

It is worth mentioning that the three dimensional (3D) CCR5 structure is not yet available. However, homology-modeling of CCR5 has been performed before [9]. Therefore, in this study, the actual human CCR5 homology model was developed using the crystal structure of CXCR4 as a structural template. Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors. This study should serve as a powerful tool in the drug design and development machinery.

2. Computational Methods

2.1. Homology Modeling of CCR5

In order for our molecular docking study to be executed, the crystal structure of human CCR5 was homology modeled using the human CCR5 protein sequence retrieved from the UniProt database [22] (Uniprot ID: P51681). The actual homology model of CCR5 was developed using the crystal structure of CXCR4 (PDB ID: 3ODU) as a structural template and using the Modeler software [23] add-on in Chimera [24]. Hydrogen atoms were included in our enzyme model, whilst all other important active site residues were identified using Chimera Multi-align Viewer [24].

2.2. Maraviroc Structure Acquisition and Preparation

Maraviroc, the known FDA-approved CCR5 antagonist, was obtained in a mol2 file format from the DrugBank [25–27]. This CCR5 antagonist had its geometry optimized and energy minimized using the MMFF94 force field found in Avogardros software [28]. Thereafter, for subsequent analyses, maraviroc was kept in the mol2 format.

2.3. Ligand Library Generation

2.3.1. Structural Similarity-Based Compound Library Generation

Maraviroc was used as the template for generating the 2D shape similarity-based compounds library from the Zinc Database. The maraviroc structure was drawn using the MarvinSketch software [29]. This reference template was used uploaded and queried the Zinc Database for all structures that had greater than 60% shape similarity to maraviroc. The query search generated a total of 1,002 compound

hits. As explained in our earlier report, certain physiochemical filters were implemented to enhance the structural query process [3]. In this study, the default physiochemical filter was set at drug-like qualities. Only compounds with molecular weight between 150 and 500 kDa were selected—this resulted in 480 hits. Other criteria were imposed to ensure the inclusion of the maximum number of compounds, such as the compounds had to have an xlog P between −4 and 5, a net charge between −5 and 5, rotatable bonds between 0 and 8, a polar surface area of between 0 and 150, have hydrogen bond donors/acceptors between 0 and 10, and polar desolvation between 0 and 1 kcal/mol whereas compounds must have an apolar desolvation between −100 and 40 kcal/mol. Thereafter, these compounds were downloaded as a single mol2 file format and were individually separated into mol2 files using the Molegro Molecular Viewer (MMV) software suite [29]. These files were then converted into a pdbqt format using the built-in Autodock Vina feature in the Raccoon software [30].

2.3.2. Pharmacophore-Based Library Generation

The pharmacophore-based compound libraries were generated using the chosen pharmacophoric regions as illustrated in Figure 2 and the ZincPharmer Database [31]. Also, keeping in mind that the desired compounds would be selected based on their probability of forming good interactions with the receptor. Furthermore, the Lipinski Rule of Five was imposed as the set criterion for screening compounds not only for confining conformational variations of the same ligand, but also for reducing any duplication The query search generated a total of 602 compound hits. All these compounds were downloaded as a single sdf file format and then separated and processed as explained above in Section 2.3.1.

Figure 2. Maraviroc structure used as a template for pharmacophore-based and structural similarity-based compound library generation. Pharmacophore selection criteria—green depicts hydrophobicity, purple depicts aromatic and white depicts hydrogen donor. Arrows indicate that constraints have been imposed.
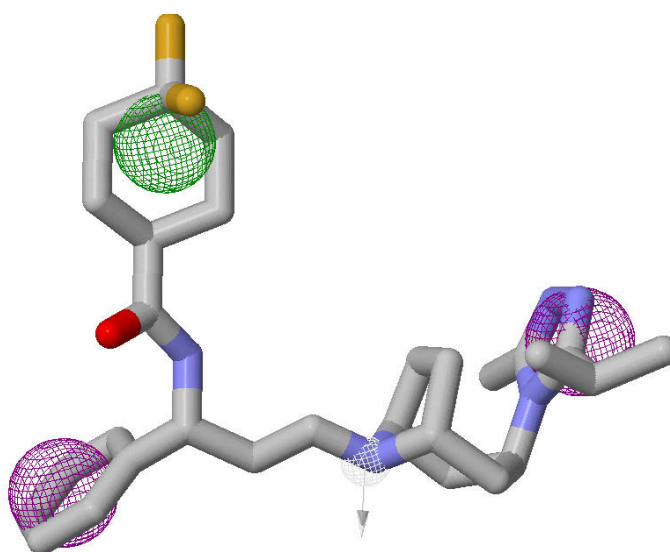
2.4. Virtual Screening and Validation of Docking Protocol

In our study, the known CCR5 antagonist (maraviroc) and the two respective ligand-based compound libraries were subjected to virtual screening against the CCR5 enzyme. The Autodock Vina [32] screening

software was used to conduct docking calculations. Although the Screening procedure was run using the software default settings; the exhaustiveness of the screening was fixed to the value of 8. The grid box used to define the screening site was verified by using the built-in functionality property found in Autodock Vina [24]. The grid box was defined around the following key amino acid residues, namely Phe85, Trp86, Trp 89, Leu104, Tyr108, Ile 198, Try251 and Glu283,and these resembled the active site residues found in the crystal structure of CXCR4 enzyme following the sequence alignment performed in Chimera. The X, Y and Z centres were defined as 11.01, −2.08 and 45.68, whereas the X, Y and Z size dimensions were defined as 58, 82 and 74, respectively. Autodock Vina screening results were produced in the pdbqt format. From each of the two compound libraries, the top ten compounds were selected on the basis of best binding affinities and visualized using the Viewdock feature in Chimera.

2.5. Molecular Dynamics Simulations and Post-Dynamic Analysis

The best-docked ligand-enzyme complexes that resulted from the structure-based and pharmacophore-based compounds library were then exposed to MD stimulations using the Amber software [33], following the procedure explained in our previous report [34] were performed using the We examined the post-dynamic nature of how ligands interacted with the CCR5 target protein within a range of 5 Å as illustrated by hydrogen bond and hydrophobic interactions using the Molecular Viewing Operator (MOE) program [35]. Likewise, residue contribution towards ligand binding was computed using the Moldock scoring functions [29].

2.6. Three-Dimensional (3D) QSAR Analysis

A set of 35 novel oxamino-piperidino-piperidine amide analogs (Figure 3) with available IC_{50} (mM) data was taken from literature for the development of the atom-based 3D-QSAR model (Table 1) [1]. This 3D-QSAR study was performed in Discovery studio 3.5 [36]. The $1/\log IC_{50}$ value of CCR5 was used in this study. Of the 35 compounds reported, 26 compounds were used as a training set and the remaining nine compounds were used as a test set, based on a random selection. The compounds in the test set have a range of biological activity values similar to that of the training set. The ligands were pre-aligned using a molecular overlay method and placed in a 3D grid space (Figure 4). The grid spacing was 1 Å. The energy potentials on every grid point were then calculated using a CHARMm force field which used the electrostatic potential and the Van der Waals potential and treated as separate terms. A +1e point charge is used as the electrostatic potential probe and distance-dependent dielectric constant is used to mimic the solvation effect. For the Van der Waals potential a carbon atom with a 1.73 Å radius is used as a probe. The energy grid potentials can be used as independent variables to create partial least-squares (PLS). Furthermore, the best 3D-QSAR model was validated by predicting activities of the 9 test set compounds. The 3D-QSAR was evaluated by cross-validated R^2, Q^2. The predicted $1/\log IC_{50}$ at 6th PLS factor are tabulated in Table 1.

Figure 3. The 2D structures for the oxamino-piperidino-piperidine amide analogs used in the 3D QSAR of this work.
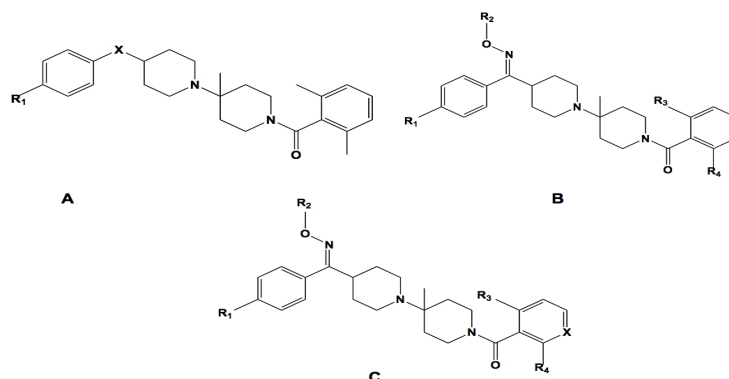

Table 1. Dataset analyzed for 3D QSAR with experimental $1/\log IC_{50}$, predicted $1/\log IC_{50}$ and residual value.

#	Core	X	R_1	R_2	R_3	R_4	Expt. $1/\log IC_{50}$ (mM)	Prdt. $1/\log IC_{50}$ (mM)	Residual
1	A	CH_2	Br	-	-	-	0.250	0.259	0.009
2	A	NH_2	Br	-	-	-	0.267	0.265	0.002
3	A(E)	$=N–OCH_3$	Br	-	-	-	0.290	0.270	0.020
4	A(E)	$=N–OCH_3$	Br	-	-	-	0.252	0.271	−0.019
5	B	-	Br	CH_3	Cl	NH_2	0.360	0.359	0.001
6	B	-	Br	CH_3	CH_3	OH	0.278	0.276	0.001
7	C	$N^+–O$	CH_3	CH_3	CH_3	CH_3	0.267	0.286	−0.018
8	C	N	Cl	C_2H_5	CH_3	CH_3	0.435	0.392	0.043
9	A	CH_2	Cl	-	-	-	0.229	−0.248	−0.018
10	A	CH_2	I	-	-	-	0.253	0.230	0.023
11	A	CH_2	CF_3	-	-	-	0.333	0.269	0.064
12	A	CH_2	CH3	-	-	-	0.218	0.233	−0.015
13	A	CH_2	OCH_3	-	-	-	0.227	0.233	−0.006
14	A	CH_2	SO_2CH_3	-	-	-	0.260	0.273	−0.013
15	A	$C=CH_2$	Br	-	-	-	0.333	0.313	0.020
16	B(Z)	-	Br	H	CH_3	CH_3	0.250	0.245	0.005
17	B(Z)	-	Br	C_4H_9	CH_3	CH_3	0.266	0.275	−0.009
18	B	N	Br	CH_2-CO-$NHCH_3$	CH_3	CH_3	0.274	0.292	−0.016
19	B	N	Br	C_2H_5	F	CF_3	0.301	0.305	−0.004
20	C	$N^+–O$	Br	C_2H_5	H	CH_3	0.318	0.289	0.290
21	C	$N^+–O$	Br	C_2H_5	CH_3	CH_3	0.338	0.382	−0.044
22	C	$N^+–O$	Br	C_2H_5	H	CH_3	0.310	0.307	0.003
23	C	$N^+–O$	CF_3	CH_3	CH_3	CH_3	0.329	0.324	0.005
24	C	$N^+–O$	OCF_3	CH_3	CH_3	CH_3	0.270	0.298	−0.028
25	C	$C=O$	OCF_3	C_2H_5	CH_3	CH_3	0.307	0.314	−0.007
26	C	-	Cl	C_2H_5	CH_3	CH_3	0.324	0.324	−0.006
27	A	N	Br	-	-	-	0.243	0.261	−0.018
28 [t]	B	$N^+–O$	Br	CH_3	CH_3	NH_2	0.371	0.385	0.086
29 [t]	C	CH_2	Br	CH_3	CH_3	CH_3	0.371	0.329	0.076
30 [t]	C	CH–OH	Br	C_2H_5	CH_3	CH_3	0.301	0.302	0.001
31 [t]	A	-	-	-	-	-	0.270	0.277	−0.007
32 [t]	A	-	Br	-	-	-	0.205	0.254	−0.54
33 [t]	B(Z)	$N^+ = O$	Br	C_3H_7	CH_3	CH_3	0.310	0.271	0.38
34 [t]	C		CH_3	CH_3	-	-	0.263	0.254	−0.091
35 [t]	C		CF_3	C_2H_5	CH_3	CH_3	0.321	0.306	0.015

t = test set. The conformation of compound denoted in brackets ( ) [1].

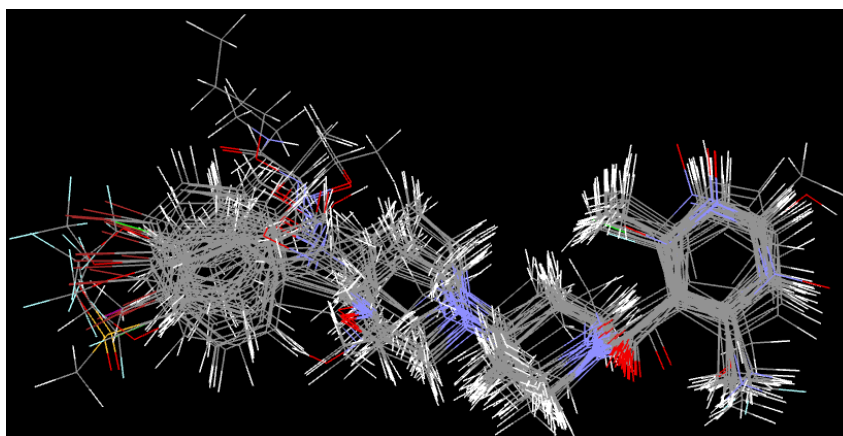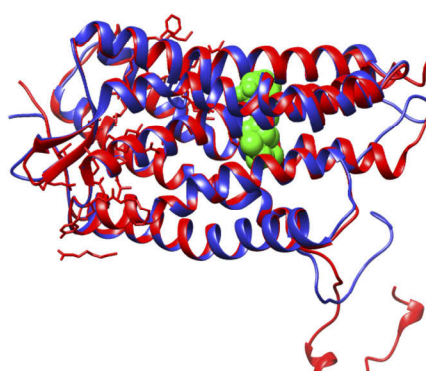**Figure 4.** Molecular alignments used in the present study.

Figure 5. (**A**) Superimposed structures of 3ODU [16] and modeled CCR5 enzyme (blue) with CCR5 antagonist, maraviroc [24]; (**B**) The 2D sequence alignment of 3ODU and the homology model generated for our study. Yellow highlighting represents α-helices and green highlighting represents β-sheets. Sequences outlined in red lack 3D crystal structure.

A

B

3. Results and Discussion

3.1. Homology Modeling of CCR5

The actual homology model of human CCR5 was modeled using the 3ODU crystal structure as a structural modeling template. As outlined in Figure 5, both these proteins demonstrated good structural similarity in and around the active site residues, with most of the residues having relatively identical locations to each other. The Multi-align Viewer tool located in Chimera recorded a 42.11% shared similarity between the two proteins' sequences; after modeling, the enzyme model had a zDOPE score of 0.91 with an RMSD of 1.1771 Å. Four differences were observed between the active site sequences of our modeled CCR5 and the 3ODU model template, which included Leu204 (3ODU) corresponding to Leu104 (CCR5), Trp85 corresponding to Trp86, and two residue gaps namely corresponding to Leu213 and Phe113 from 3ODU (Table 2), respectively. We assume that these noted differences have had very minimal effects in the docking study as result of the shared structural similarity between the Leu and Trp residues, respectively. Further investigation is required to verify this assumption. All non-modeled regions were removed from the active site in order to allow for emphasis on all crucial residues and their importance in the active site of CCR5.

Table 2. Comparison of the active site residues between the modeling template (3ODU) and modeled structure.

Active site residues (3ODU)	Corresponding modeled active site residues
Glu283	Glu283
Ile198	Ile198
Leu204	Leu104 #
Leu213	*
Phe85	Phe85
Phe109	Phe109
Phe113	*
Thr195	Thr195
Trp85	Trp86 #
Trp94	Trp94
Trp248	Trp248
Tyr89	Tyr89
Tyr108	Tyr108
Tyr251	Tyr251

—Differing Residue; * —Residue Gap.

3.2. Virtual Screening

Results obtained from virtual screening for the pharmacophore-based and structure-based compound libraries are shown in Table 3. Our query of the Zinc Database for compounds bearing 2D shape similarity-based identity to the reference drug template generated 220 compound hits. However, our query of the ZincPharmer Database for pharmacologically related compounds generated 120 compound hits. All compounds in the two generated compounds libraries (see Methods sections for details) were then docked into the active site of the CCR5 enzyme using Autodock Vina and thereafter, from each library, we selected the top 10 compound hit leads. As shown in Table 3 and

Figure 6, all top 10 ranked compounds from each library exhibited remarkably higher binding energies compared to Maraviroc (−10.2 kcal/mol), with binding energies ranging from −12.2 to −11.6 kcal/mol for the 2D shape similarity-based identity library, whereas the pharmacophore-based library had binding energies ranging from −12.0 to −11.4 kcal/mol.

As outlined in Figure 6, an unexpected observation revealed that compounds that were structurally similar had higher binding affinities on average compared to those compounds that were pharmacophorically-based. Although, not a huge variance amongst both libraries existed, however, this was an indicator about the particular importance of each specific pharmacophoric area required for CCR5 antagonistic behavior. This observation might be attributed to the selected pharmacophore-based groups. A further investigation is required for the imperative role played by these pharmacophoric groups in terms of site-specific interactions and in return, how do these interactions affect inhibitor binding affinities and functioning. It was also worth mentioning that compounds 3, 5, 7 and 8 from both the structure-based and pharmacophore-based generated libraries demonstrated exact binding energies (Figure 6). Another interesting observation, not a large difference existed amongst the top 10 binding energies compounds from both the structure-based and pharmacophore-based generated libraries with a difference of −0.60 kcal/mol between the highest and lowest ranked compounds (Figure 6). This observation might be attributed to the conservancy of the crucial pharmacophoric and architecturally shared properties "amongst these compounds allowing for "alleged stability", and hence for well-maintained binding affinities.

With the intention of finding the best compounds of this study, we integrated both the generated compound libraries and revealed the top 10 best-docked compounds (Figure 7 and Table 3). It was instantly observed that these compounds were remarkably larger in size as compared to maraviroc (Figure 7). Furthermore, majority of the compounds occupied the spaces between Tyr89, Trp94, Glu283, Leu104, Thr195, Tyr251, Phe109, Ile198 and Trp248 respectively, something which maraviroc did not achieve as a result of its smaller size (Figure 7). In this study, integration of these factors might be the contributing factors for the higher binding affinities as the total number of interactions with the active site was remarkably higher compared to those felt by maraviroc.

From our docking calculations, it appeared that the virtual screening compound hits demonstrated good activity as CCR5 antagonists. Several factors might have contributed to these findings, however one of such importance to us was that the docking protocol implemented might have not been accurate enough to provide precise estimates of the different binding energies. In an effort to eradicate this factor, we opted to validate the docking method applied in this study. "Cross validation" was not employed to validate our docking results. This is an approach, where other docking programs are used to validate the data attained from the original docking software. Due to previous experience of using different docking software with various scoring functions would generate results that could be different and be misleading, we opted to disregard this docking validation approach. We strongly believe that the utmost rational mannerism for validation of docking calculations, or even any other computational tool, is to perform the calculations on a set of compounds with available experimental data and these results are then compare against known experimental data for validation.

Table 3. List of the top 10 screened compounds based on their docked binding energy. Compounds are ranked in order of highest to lowest binding affinity.
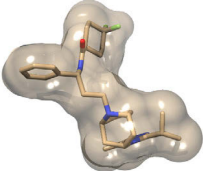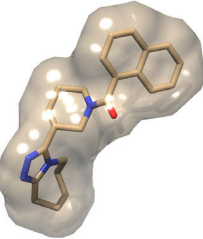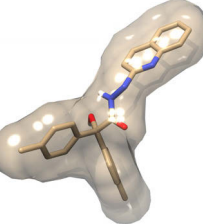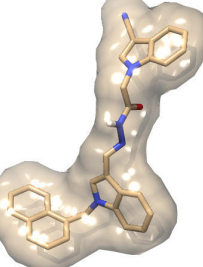
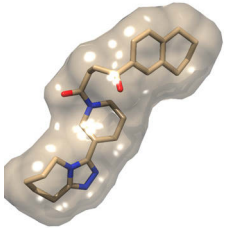| Library	Rank	ZINC ID	Structure	Binding Energy (kcal/mol)	xlogP	H-bond Donors	H-bond Acceptors	Molecular Weight (g/mol)
Ref	R	ZINC03817234		−10.2	−3.50	2	6	514.69
S *	1	ZINC71849549		−12.2	2.27	2	6	318.89
P **	2	ZINC00825224		−12.0	4.11	3	5	397.488
P	3	ZINC00634884		−12.0	5.96	1	6	481.60

Table 3. *Cont.*
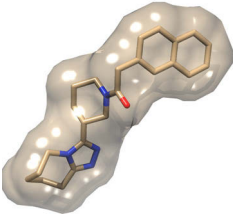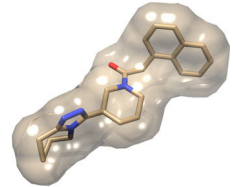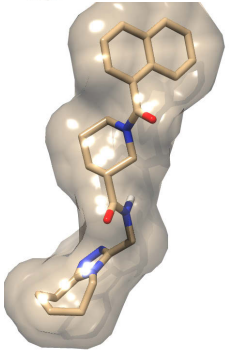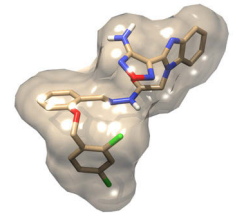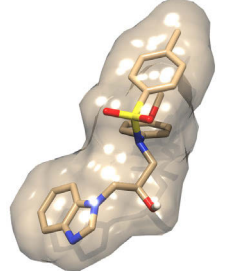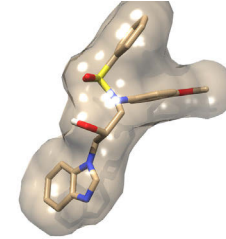
Library	Rank	ZINC ID	Structure	Binding Energy (kcal/mol)	xlogP	H-bond Donors	H-bond Acceptors	Molecular Weight (g/mol)
S	**4**	ZINC32760563		−11.9	3.47	0	5	388.51
S	**5**	ZINC32760533		−11.8	3.44	0	5	388.52
S	**6**	ZINC25010434		−11.8	2.16	1	7	431.54
P	**7**	ZINC00851466		−11.8	5.52	3	10	536.38

Table 3. *Cont.*

| Library	Rank	ZINC ID	Structure	Binding energy (kcal/mol)	xlogP	H-bond Donors	H-bond Acceptors	Molecular weight (g/mol)
S	**8**	ZINC71818945		−11.7	3.35	0	6	434.58
P	**9**	ZINC00895646		−11.7	3.99	2	7	451.55
P	**10**	ZINC00895634		−11.7	3.54	2	7	438.53

* Similarity-based library, ****** Pharmacophore-based library, Ref-Maraviroc.

Figure 6. The top 10 ranked ZINC compounds from both the 2D similarity-based and pharmacophore-based libraries.
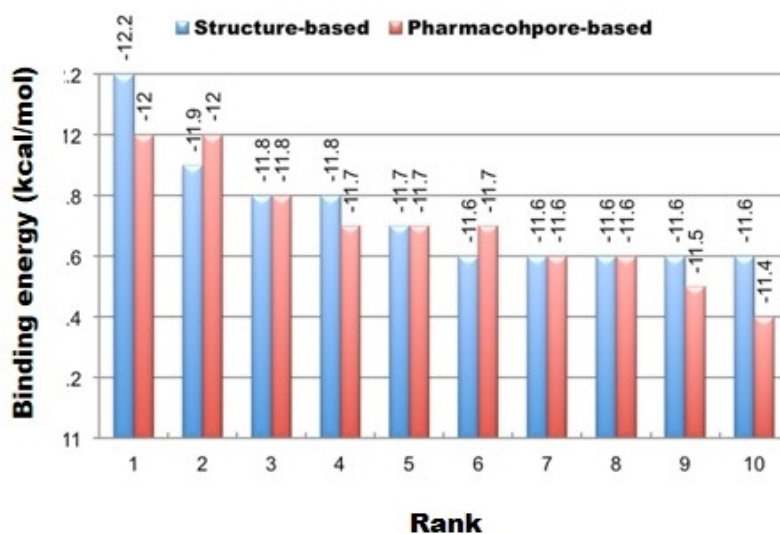

Rank	Structure-Based	Pharmacophore-Based
1	ZINC71849459	ZINC00634884
2	ZINC32760563	ZINC00825224
3	ZINC25010434	ZINC00851466
4	ZINC32760533	ZINC00877991
5	ZINC71818945	ZINC00895634
6	ZINC11546940	ZINC00895646
7	ZINC19698213	ZINC00895774
8	ZINC25010439	ZINC01038388
9	ZINC71849397	ZINC01080881

Figure 7. Docking conformations of the known CCR5 antagonist (maraviroc) and the top 10 ranked docked compounds from both the pharmacophore-based and structure-based libraries determined in this study, all-complexed with the CCR5 enzyme.
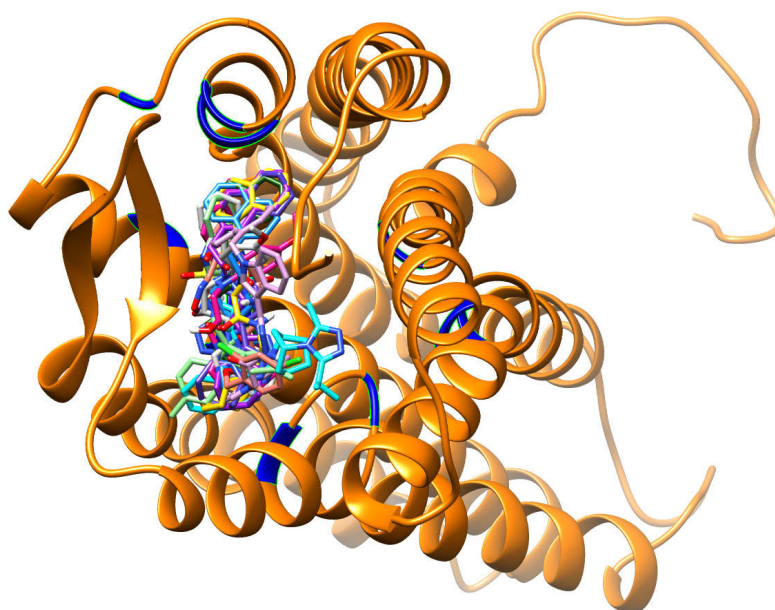
To this end, in order to validate our docking approach implemented in this study, we performed docking analyses on a set of compounds assayed using our docking method. The structures and experimental IC_{50} values were obtained from Binding database [37]. The binding energy of each compound was then plotted against its corresponding experimental IC_{50} value (Figure 8). As evident from the docking results (Figure 8), the docked energies are in great accordance with the experimental IC_{50}. We observed that as the binding affinity increased (lower binding energy), the IC_{50} increased (Figure 8). We postulate that the larger the binding affinity, the more concentration is needed for complete enzyme inhibition. To us, this proves to be an interesting trend, since the top 10 ranked compounds which were obtained from both libraries (Table 3) had higher binding energies compared to any of those used in the assay. This implies that the docking approaches used in this work could be reliable enough to estimate the binding affinities for the top 10 ranked compounds (from both libraries—Table 3).

Figure 8. The binding energies determined in our study and were compared against IC_{50} values for the compounds assayed. The higher the binding affinity, the higher the IC_{50}.

|	1	2	3	4	5
Binding energy	-11.8	-10.9	-9.1	-8.1	-7.6
IC50	0.2	2.1	10	23	26

Inhibitor
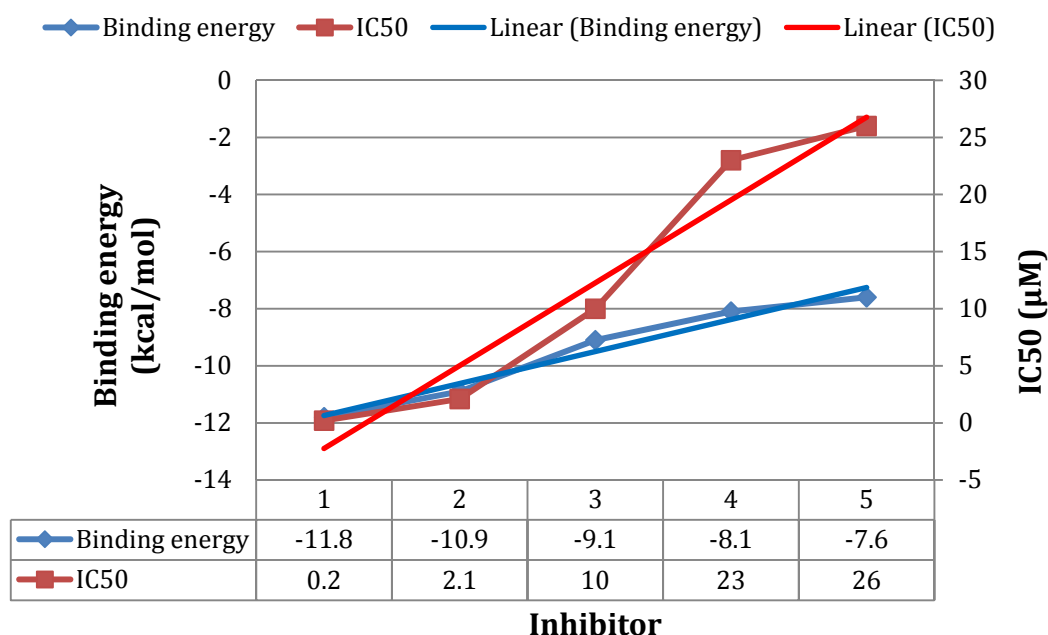
3.3. Molecular Dynamics Simulations

In order to gain more insight into the stability of the resulted highest ranked virtual screening hit complexes, the nature of overall interaction theme between the proposed ligands and the target protein and the specific amino acids involved in ligand binding, we performed 1 ns MD stimulations followed by extensive post-dynamic analyses on the ligand-enzyme complexes.

MD stimulations of 1 ns were performed for the highest ranked virtual screening hit complexes (Figure 9), to ensure the stability of the ligand within the CCR5 active site. We also performed MD simulations on the reference ligand (maraviroc) bound to CCR5 (see Supplementary Material). From our previous experience with molecular docking, in many occasions, we experienced that even best docked structures may fly away from the enzyme active site within a few picoseconds of

MD stimulations. Therefore, we believe that docking calculations that are not validated by relatively a long MD run to ensure stability of the system might not be reliable. Interestingly, for all the compound-enzyme complexes, the average RMSD values were below 2.5 Å. In addition, the variability of the potential energies fell with 1000 kcal/mol and this suggested being a good indicator of the system stability.

Figure 9. The highest ranked virtual screening hit lead complexes with CCR5 subjected to MD simulations. Structure-based compound (ZINC71849459) in complex with CCR5 (Black). Pharmacophore-based compound (ZINC00634884) with CCR5 (Blue).
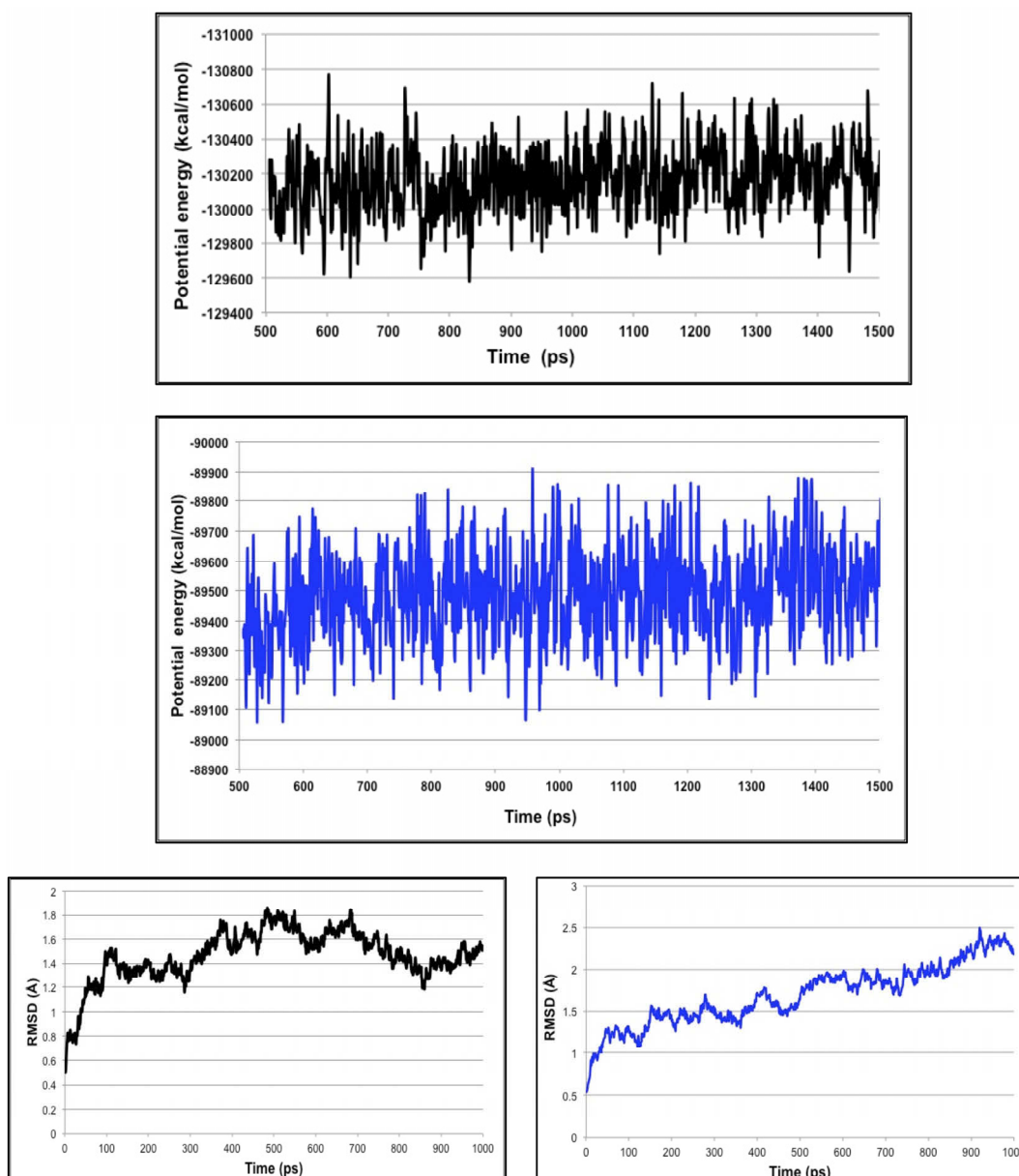

3.4. Per-Residue Interactions

In an effort to investigate the contribution of a single amino acid towards ligand (and/or antagonist) binding, we computed per-residue interactions using Moldock software [29] (Figure 10). The top ranked antagonist with the highest binding energy within each compound library was assessed, which

included ZINC71849459 from the structure-based identity library and ZINC00634884 from the pharmacophore-based library. We noticed that Lys197, Phe109, Trp86, Tyr89 and Try108 exhibited remarkably interactions with both of those docked ligands. However, Trp86, Tyr89, Thr105, and Tyr108 demonstrated especially better interactions with ZINC00634884 (pharmacophore-based library), whereas Lys197, Leu196 and Glu283 showed good interactions with ZINC71849459 (structure-based identity library). As shown in Figure 11, in order to gain a better understanding of the ligand-amino acid interactions occurring within CCR5's active site, we generated a plot of the specific amino acid residues-ligand interactions using the MOE software [35].

Figure 10. Per-residue interactions for the highest ranked compounds with the best binding energy from the structure-based and pharmacophore-based libraries.
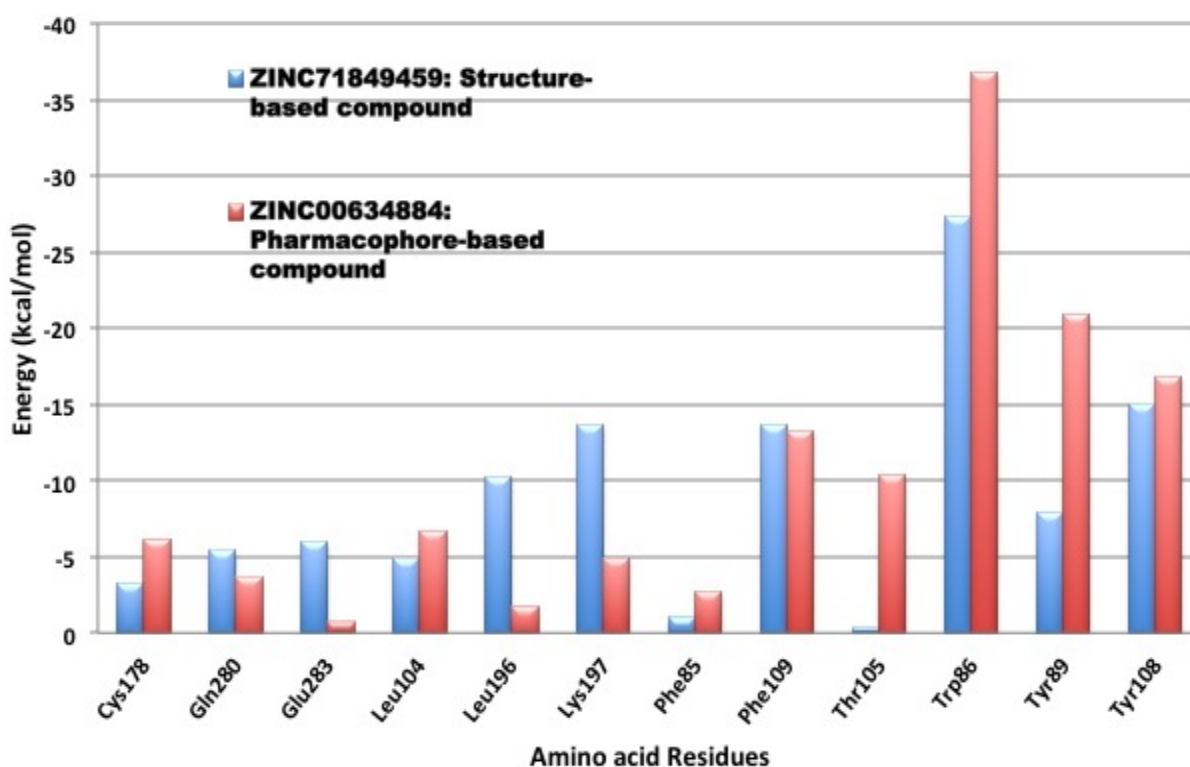
The MOE plot analysis of ZINC71849459 bound to CCR5 enzyme's active site revealed that the ligand was especially well surrounded electrostatically by several amino acid residues within the active site (Figure 11A). However, it could be noted that the ligand ZINC00634884 bound to the active site was not well cradled electrostatically by the amino acid residues compared to the former (Figure 11B). An interesting observation was that not a single amino acid formed any hydrogen bonds between the protein and ligand for anyone of the two ligands investigated. As shown in Figure 11, on average the hydrophobic amino acid residues played crucial roles in protein-ligand interactions than the hydrophilic amino acid residues. Another important observation was that the stronger ligand-amino acid residue interactions exhibited here in Figure 11 matched to those outlined in Figure 10.

Figure 11. Pharmacophore-based compound (ZINC00634884) with CCR5 (**A**) and Structure-based compound (ZINC71849459) in complex with CCR5 (**B**), respectively, showing the hydrogen bonding and electrostatic interactions with the enzyme's active site using MOE.

(**A**)

(**B**)
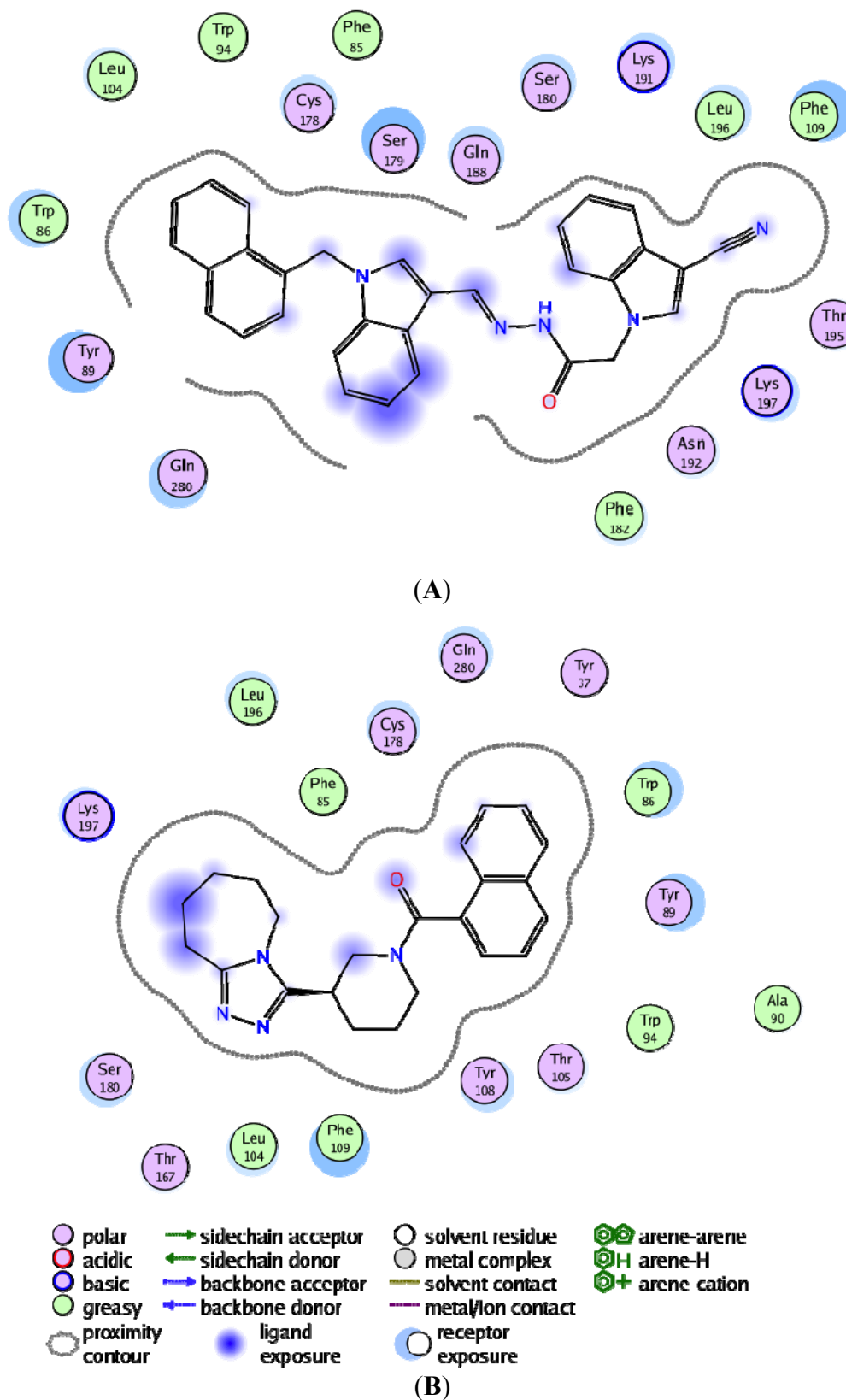
3.5. Atom-Based 3D-QSAR

The atom-based 3D-QSAR model was developed from the training set of 26 inhibitors (Table 1) and the test set of 9 inhibitors using molecular overlay alignments (Figure 3). This atom-based 3D-QSAR model was built after model development and validation based on the internal predictions of the training set and the external predictions of the test set. PLS analyses of the CCR5 inhibitor training sets showed a high cross-validated r_{cv}^2 value of 0.84 using three principal components and non-cross-validated r^2 value of 0.941. All of the parameters of these QSAR model showed certain reliability and feasible predictability to help us design new and high selectivity CCR5 inhibitors. From Figure 12, we can see that almost all compounds in the test set and training set yielded a good predicted value. The graphical plot of observed *vs.* calculated TPH1 inhibitory activity for both the training set as well as the test set is shown in Figure 12.

Figure 12. Correlation graph between the experimental 1/log IC50 and predicted 1/logIC50.
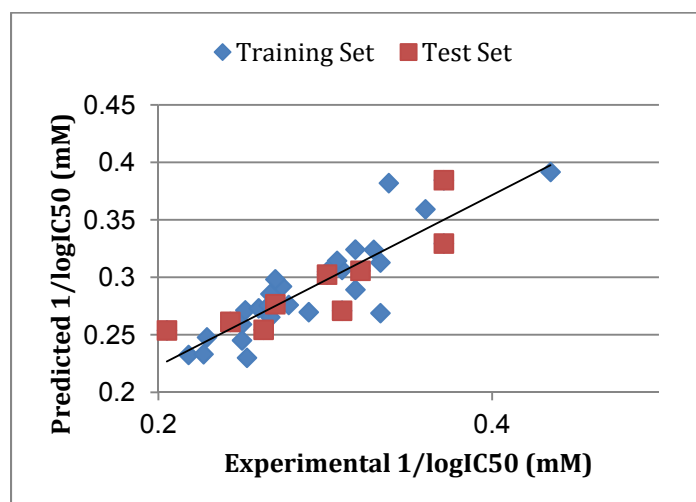
4. Conclusions

In the present work, the structure of human CCR5 was homology modeled using the crystal structure of CXCR4 as a structural modeling template. The protein appeared to be modeled with a remarkably degree of accuracy, specifically at the active site where docking studies were performed. Our query of the ZINC database for drug-like compounds that shared 2D shape similarity-based identity and query of ZINCPharmer for pharmacologically related drug-like compounds to a known FDA-approved CCR5 antagonist called maraviroc. The entire top 10 ranked compounds from each library exhibited remarkably higher binding energies compared to the best-docked structure of maraviroc. Moreover, an unforeseen observation revealed that compounds that were architecturally similar had higher binding energies on average compared to those that were pharmacophorically-based. To validate our docking calculations, the same docking approach adopted for the ligand-based libraries was performed on a set of compounds with known experimental data attained from inhibition assays against HIV-1 CCR5 and the results were compared against experimental data. The docked energies are in great accordance with the experimental IC$_{50}$. It was concluded that compounds with more

favorable predicted binding energies than the known drug maraviroc will have "better activity" (*i.e.*, smaller IC_{50} values). To us, this proves to be an interesting trend, since the top 10 ranked compounds which were obtained from both libraries had higher binding energies compared to any of those used in the biological assay, therefore it stands to reason compound hits elucidated in this study exhibit humble activity as CCR5 antagonists. Also, this implies that the docking approaches used in this work could be reliable enough to estimate the binding affinities for other compounds to be studied. Furthermore, from our previous experience with molecular docking, and in numerous insistences, the reliability of a stable protein-ligand complex might not be a true reflection. Therefore, in order to obtain more insight on the stability of the resulted docked complexes, the nature of the overall interaction themes between the generated ligands and the target protein, and the specific amino acids involved in the ligand binding, we performed 5 ns MD simulations followed by extensive post-dynamic analyses on the ligand-enzyme complexes resulted from our docking simulations. We took our study a step further by obtaining a set of 35 novel oxamino-piperidino-piperidine amide analogs with available IC_{50} (mM) data taken from literature for the development of our atom-based 3D-QSAR model. All of the parameters of the QSAR model showed certain reliability and feasible predictability to help us design new and high selectivity CCR5 inhibitors. Our novel identified leads have the propensity to be considered as potential CCR5 antagonists and moreover as potential HIV-1 entry inhibitors. Information gained from this study could shed light on the activity of a new series of lead compounds as potential HIV entry inhibitors and should serve as a powerful tool in the drug design and development machinery.

Supplementary Materials

Supplementary materials can be accessed at: http://www.mdpi.com/1420-3049/19/4/5243/s1.

Acknowledgments

Author Contributions

S.M.: Responsible for all of the computational experimental work and data analyses as well as was responsible for the entire manuscript write up from the cover page to the references; R.C.D.: was only responsible for the 3D-QSAR computational aspect which included analysis of results and write up; M.E.S.: principal investigator and corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Gadhe, C.G.; Kothandan, G.; Cho, S.J. Computational modeling of human coreceptor CCR5 antagonist as a HIV-1 entry inhibitor: Using an integrated homology modeling, docking, and membrane molecular dynamics simulation analysis approach. *J. Biomol. Struct. Dyn.* **2013**, *31*, 1251–1279.

2. Xu, Y.; Liu, H.; Niu, C.Y.; Luo, C.; Luo, X.M.; Shen, J.H.; Chen, K.X.; Jiang, H.L. Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-yl)-butanes based on the structural modeling of human CCR5 receptor. *Bioorgan. Med. Chem.* **2004**, *12*, 6193–6208.

3. Soliman, M.E.S. A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors. *Drug Dev. Res.* **2013**, *74*, 283–295.

4. Johnson, B.C.; Pauly, G.T.; Rai, G.; Patel, D.; Bauman, J.D.; Baker, H.L.; Das, K.; Schneider, J.P.; Maloney, D.J.; Arnold, E.; *et al.* A comparison of the ability of rilpivirine (TMC278) and selected analogues to inhibit clinically relevant HIV-1 reverse transcriptase mutants. *Retrovirology* **2012**, *9*, 1–23.

5. Patel, J.R.; Prajapati, L.M. Predictive QSAR modeling on tetrahydropyrimidine-2-one derivatives as HIV-1 protease enzyme inhibitors. *Med. Chem. Res.* **2013**, *22*, 2795–2801.

6. Zhan, P.; Chen, X.; Li, D.; Fang, Z.; de Clercq, E.; Liu, X. HIV-1 NNRTIs: Structural diversity, pharmacophore similarity, and impliations for drug design. *Med. Res. Rev.* **2013**, *33*, E1–E72.

7. Johnson, B.C.; Metifiot, M.; Ferris, A.; Pommier, Y.; Hughes, S.H. A Homology Model of HIV-1 Integrase and Analysis of Mutations Designed to Test the Model. *J. Mol. Biol.* **2013**, *425*, 2133–2146.

8. Pani, A.; Loi, A.G.; Mura, M.; Marceddu, T.; la Colla, P.; Marongiu, M.E. Targeting HIV: Old and new players. *Curr. Drug Targets* **2002**, *2*, 17–32.

9. Fano, A.; Ritchie, D.W.; Carrieri, A. Modeling the structural basis of human CCR5 chemokine receptor function: From homology model building and molecular dynamics validation to agonist and antagonist docking. *J. Chem. Inf. Model.* **2006**, *46*, 1223–1235.

10. Manikandan, S.; Malik, B.K. Modeling of human CCR5 as target for HIV-I and virtual screening with marine therapeutic compounds. *Bioinformation* **2008**, *3*, 89–94.

11. Cormier, E.G.; Persuh, M.; Thompson, D.A.D.; Lin, S.W.; Sakmar, T.P.; Olson, W.C.; Dragic, T. Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5762–5767.

12. Farzan, M.; Vasilieva, N.; Schnitzler, C.E.; Chung, S.; Robinson, J.; Gerard, N.P.; Gerard, C.; Choe, H.; Sodroski, J. A tyrosine-sulfated peptide based on the N terminus of CCR5 interacts with a CD4-enhanced epitope of the HIV-1 gp120 envelope glycoprotein and inhibits HIV-1 entry. *J. Biol. Chem.* **2000**, *275*, 33516–33521.

13. Cormier, E.G.; Tran, D.N.; Yukhayeva, L.; Olson, W.C.; Dragic, T. Mapping the determinants of the CCR5 amino-terminal sulfopeptide interaction with soluble human immunodeficiency virus type 1 gp120-CD4 complexes. *J. Virol.* **2001**, *75*, 5441–5449.

14. Cocchi, F.; DEVico, A.L.; Garzino-Demo, A.; Arya, S.K.; Gallo, R.C.; Lusso, P. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **1995**, *270*, 1811–1815.

15. Pease, J.; Horuk, R. Chemokine receptor antagonists. *J. Med. Chem.* **2012**, *55*, 9363–9392.

16. Ditzel, H.J.; Rosenkilde, M.M.; Garred, P.; Wang, M.; Koefoed, K.; Pedersen, C.; Burton, D.R.; Schwartz, T.W. The CCR5 receptor acts as an alloantigen in CCR5Delta32 homozygous individuals: Identification of chemokineand HIV-1-blocking human antibodies. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5241–5245.

17. Kothandan, G.; Gadhe, C.G.; Cho, S.J. Structural Insights from Binding Poses of CCR2 and CCR5 with Clinically Important Antagonists: A Combined In Silico Study. *PLoS One* **2012**, *7*, e32864.

18. Perez-Nueno, V.I.; Ritchie, D.W.; Rabal, O.; Pascual, R.; Borrell, J.I.; Teixido, J. Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48*, 509–533.

19. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P.A.; Markopoulos, J.; Igglessi-Markopoulou, O. Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques. *J. Comput. Aided Mol. Des.* **2006**, *20*, 83–95.

20. Aher, Y.D.; Agrawal, A.; Bharatam, P.V.; Garg, P. 3D-QSAR studies of substituted 1-(3,3-diphenylpropyl)-piperidinyl amides and ureas as CCR5 receptor antagonists. *J. Mol. Model.* **2007**, *13*, 519–529.

21. Kellenberger, E.; Springael, J.-Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J.-L.; Rognan, D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.* **2007**, *50*, 1294–1303.

22. Consortium, T. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. **2012**, *40*, D71–D75.

23. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **2007**, doi:10.1002/0471140864.

24. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

25. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu V.; *et al.* DrugBank 3.0: A comprehensive resource for "Omics" research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035–D1041.

26. Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906.

27. Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.

28. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminf.* **2012**, *4*, doi:10.1186/1758-2946-4-17.

29. Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 5.12.1 (version 5), 2013, ChemAxon (http://www.chemaxon.com).

30. Thomsen, R.; Christensen, M.H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.

31. Forli, S. AutoDock|Raccoon: An automated tool for preparing AutoDock virtual screenings, 2013.

32. Koes, D.R.; Camacho, C.J. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* **2012**, *40*, W409–W414.

33. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

34. Ahmed, S.M.; Kruger, H.G.; Govender, T.; Maguire, G.E.M.; Sayed, Y.; Ibrahim, M.A.A.; Naicker, P.; Soliman, M.E.S. 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.* **2013**, *81*, 208–218.

35. The Molecular Operating Environment (MOE) available under license from Chemical Computing Group Inc., 1010 Sherbrooke St. W. Suite 910, Montreal, Quebec, Canada H3A 2R7.

36. *Discovery Studio Modeling Environment*, Release 3.5; Accelrys Software Inc.: San Diego, CA, USA, 2012.

37. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucl. Acids Res*. **2007**, *35*, D198–D201.

Sample Availability: Not available.