UNIVERSITY OF KWAZULU-NATAL

COMPUTATIONAL STUDIES OF PENTACYCLOUNDECANE PEPTIDE BASED HIV-1 PROTEASE INHIBITORS

2013

BUYISILE CHIBI

COMPUTATIONAL STUDIES OF PENTACYCLOUNDECANE PEPTIDE BASED HIV-1 PROTEASE INHIBITORS

BUYISILE CHIBI

2013

A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal, Westville, for the degree of Master's in medical sciences.

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

Supervisor:
Signed:---------- Name: **M. E. S. Soliman** Date:----------

Co-Supervisor:
Signed:---------- Name: **H. G. Kruger** Date:----------

Co-Supervisor:
Signed:---------- Name: **T. Govender** Date:----------

Co-Supervisor:
Signed:---------- Name: **G. E. M. Maguire** Date:----------

Abstract

HIV/AIDS is a pandemic disease infecting millions of people worldwide. The circumstance and life expectancy of HIV-infected patients have changed to a manageable condition with the aid of antiviral therapy. Amongst all the HIV infections in South Africa, approximately 95% are of subtype C strain. The core aim of this study was first, to investigate the flap dynamics of the C-SA HIV-1 PR in its free (apo enzyme) and bound forms by applying molecular dynamics (MD) simulations, and second to develop an effective computational model for the identification of potential novel PCU-based peptide HIV protease inhibitors, employing a structure-based virtual screening approach.

HIV protease consists of two essential flap tips whose conformations and dynamics are crucially important for substrate entry to access the protease catalytic site for its mechanism. There are few studies on MD simulations of C-SA HIV-1 PR for the investigation of enzyme flexibility. MD simulations on the free enzyme as well as the docked PR complex of five previously reported PCU-based peptides were studied. The dynamic behaviour of the free enzyme and the inhibitor-PR complex were compared to that of the ritonavir-enzyme complex. An implicit solvation model was used and the MD simulation was performed over 10 ns. The overall results of the free enzyme calculations showed that C-SA HIV-1 PR has greater flexibility in the flaps in comparison to subtype B. For C-SA PR, the flap opening was observed at 3.5 ns and a fully-opened form of the protease at 8.6 ns. During the 10 ns simulation, flap opening and closing of the C-SA protease was observed which confirms that suitable equilibration was reached and that the opening was not an artefact. On the other hand, the subtype B PR was observed to open at 3.4 ns and remained opened for the entire simulation.

MD simulations of the PR complexes with the five PCU-peptide inhibitors as well as ritonavir showed that all complexes attained a closed conformation. In this study, four metrics have been used to estimate the flap flexibility: (1) the distance between the flap tips (Cα I50 - Cα I149), (2) the distance between the α-carbons of K55 and K154 residues, (3) the distance between the flap tips and catalytic residues (Cα I50 - Cβ D25) and (Cα I149 – Cβ D124) were used to measure the extent of flap opening and (4) the

angle formed from the α-carbons of residues G48-G49-I50 of monomer A and G147-G148-I149 of monomer B were measured to monitor the flap-curling behaviour. The first metric has often been used in most reported studies. It was observed that the first metric takes flap-curling effect into account which can lead to erroneous interpretation of flap openings from the MD results. The second metric is insensitive to flap-curling and movements observed correspond to real flap openings. When the movement observed using the third metric, asymmetric behaviour of the flaps was observed due to the interaction of the inhibitor with the flaps and the catalytic site. These interactions result in the shrinking of the active site thus, holding the inhibitor more tightly. The four metrics do not clearly define the conformation of the flaps or the extent of flap opening. In this study, the second metric (Cα K55 - Cα K154) seemed to be more appropriate in estimating the extent of flap opening.

This thesis also, involves the virtual screening of 640 PCU-based peptide compounds (R-Cage-R, with R = *(R)*- or *(S)*-amino acids) as potential HIV PR inhibitors. This study was an attempt to gain insight into the possible structure-activity relationship (SAR) of these potential inhibitors against C-SA HIV-1 PR. The pentacycloundecane (PCU) cage diol were either functionalised as HOOC-$(CH_2)_n$-Cage-$(CH_2)_n$-COOH or H_2N-$(CH_2)_n$-Cage-$(CH_2)_n$-NH_2 with n = 0, 1, 2 or 3. Virtual attachment of a mixture of *(R)*- and *(S)*-amino acids to the cage result in four possible diastereomeric products: the *(R,R)-, (R,S), (S,R)*- and *(S,S)* forms of R-Cage-R, with R = *(R)*- or *(S)*-amino acids. A total of 640 structures were docked into the active site of the C-SA PR and each family (acid or amide) with the sub-families (n = 0 – 3), were rank ordered according to the calculated binding energies. The future plan in this regard is to start a new project to calculate the binding free energies from MD simulations of the 640 inhibitors after which the synthetic group will attempt the synthesis and HIV PR testing of the ten best compounds.

Results obtained from this study provide deeper insight into the structural and mechanistic features of the C-SA HIV PR, which will assist in designing of potential PR inhibitors.

Declaration

I, ……., 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 section.

Signed

Acknowledgement

God's love and His Word has kept me strong throughout this work.

I would like to give thanks to:

- My supervisors: Dr Mahmoud Soliman, Prof. Gert Kruger, Dr Thavi Govender and Dr Glenn Maguire for their constant flow of support and knowledge. This has inspired and developed me as a researcher.

- My family for their unconditional love and support in all angles of life.

- My husband, Voice Mnisi for his constant support, understanding and encouraging words.

- My son, Amukelani Mnisi as a source of inspiration.

- My friends for their help and encouraging words.

- My colleagues: Bahareh Hornarparvar, Mahasin Dinat, Yushir Maharaj and Ndumiso Mhlongo for all the good and bad times we had when running jobs.

- Dr Ebrahim Mahmoud for the technical support.

- Professor Y. Sayed, HIV Proteins Research Thrust, School of Molecular and Cell Biology, University of the Witwatersrand. The source of the South African HIV-1 subtype C protease sequence and structure.

- The Center for High Performance Computing (CHPC) for the availability of cluster and helpdesk members.

- To the NRF and UKZN for financial support.

Table of contents

List of figures

List of abbreviations

3D	Three-dimensional
AIDS	Acquired Immunodeficiency Syndrome
Ala	Alanine
AM1	Austin Model 1
AM1 BCC	Austin Model 1 with Bond Charge Correction
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
AZT	Zidovudine
CA	Alpha carbon (α-carbon)
CB	Beta carbon (β-carbon)
C-SA	Subtype C South African
Cys	Cysteine
D	Aspartic acid
DNA	Deoxyribonucleic acid
FDA	USA Food and Drug Administration
GAFF	General Amber Force Field
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
HAART	Highly active antiretroviral therapy
His	Histidine
HIV	Human Immunodeficiency Virus
I	Isoleucine
IN	Integrase
K	Lysine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine

MD	Molecular Dynamics
Met	Methionine
MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics Generalized Born Surface Area
NDDO	Neglect of Diatomic Differential Overlap
NMR	Nuclear magnetic resonance
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
NtRTI	Nucleotide reverse transcriptase inhibitor
PCU	Pentacycloundecane
PDB	Protein Data Bank
PES	Potential Energy Surface
Phe	Phenylalanine
Phg	2-phenylglycine
PI	Protease inhibitor
PR	Protease
Pro	Proline
QM/MM/MD	Quantum mechanics/Molecular mechanics/Molecular dynamics
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RNA	Ribonucleic acid
SA	Surface area
SAR	Structure-activity relationship
Ser	Serine
SIV	Simian Immunodeficiency Virus
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
UNAIDS	The Joint United Nations Programme on HIV/AIDS
Val	Valine

Chapter 1

This chapter briefly explains the background of the Human Immunodeficiency Virus (HIV), a causative agent of Acquired Immunodeficiency Syndrome (AIDS)[1-4] including anti-HIV therapies employed in the past and today.

1.1 Background

AIDS is the most devastating disease outbreak to date. It is caused by a previously unknown agent,[5] which was reported in the mid- to late 1970s[6]. Since it was demonstrated that HIV is the source of AIDS, serious research efforts on developing new and improved therapies targeting HIV were initiated. Zidovudine (AZT) was synthesized in 1964 as a drug for cancer treatment but it was proven to be ineffective[6]. After the discovery that it inhibits HIV, it became the first drug to be approved by the United State Food and Drug Administration (FDA) in 1987 for HIV/AIDS treatment[7, 8]. The latest global figures from The Joint United Nations Programme on HIV/AIDS (UNAIDS) revealed that almost 2.7 million people were newly infected with HIV, 1.8 million AIDS-related deaths for both children and adults and 34 million people both adults and children were living with HIV in 2010[9].

Remarkable effort had been directed to the understanding of the HIV structure and life cycle since the spread of AIDS started to increase exponentially worldwide[5]. In nature, HIV exists in two forms, HIV-1 and HIV-2 which are closely related to the Simian Immunodeficiency Virus (SIV)[4]. Three enzymes were found to be essential in the HIV-1 life cycle; thus, these enzymes were identified as potential drug targets. These enzymes are reverse transcriptase (RT), integrase (IN) and protease (PR)[5]. This led to the development and improvement of antiviral drugs targeting different stages of the viral life cycle[5]. The mode of action of antiviral drugs is discussed in Section 1.3.

The ineffectiveness of AZT monotherapy[7, 10] and the increased number of infected patients led to the introduction of combination therapy of the anti-HIV drugs targeting different mechanisms in the HIV-1 life cycle[11]. Since the mid-1990s a combination therapy which is known as Highly Active Antiretroviral Therapy (HAART) was introduced which has changed the condition and life expectancy in HIV/AIDS patients[11-13]. This treatment includes combinations of a protease inhibitor, a nucleoside reverse transcriptase inhibitor and a non-nucleoside reverse transcriptase inhibitor[13]. With the knowledge and understanding of the transmission and prevention of the virus, the use of anti-HIV drugs have changed the effect of the virus to a manageable condition and thus, death rates of HIV patients decreased noticeably[7, 14].

On the other hand, the long term use of anti-HIV drugs results in high prevalence side effects in patients[15-17]. Effects such as lipodystrophy[7], hyperlipidemia[13], hypertotoxicity, hyperglycemia, diarrhoea, insulin resistance[13], an increase in cardiovascular disease and birth defects have been reported in HIV/AIDS patients[13, 17, 18]. Hypertotoxicity, which is a common name for liver damage, and skin rash were observed in patients taking all FDA-approved anti-HIV drugs from nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), and protease inhibitors (PIs)[19]. The use of PIs has been reported to cause hyperglycemia resulting in high levels of glucose in the blood stream. Ritonavir was reported to cause hyperlipidemia, which is an increase in the amount of fat (such as cholesterol and triglycerides) in the blood, more than for any other PIs[19]. AZT dramatically accelerates patient somatic mitochondrial DNA mutations causing a much faster aging process, especially in terms of the brain. Patients also often develop heart disease, dementia, and other age-related illnesses[20].

Other than side effects, numerous drug resistant strains have developed[21]. Due to the observed side effects researchers are now aiming to discover anti-HIV drugs that will result in a simpler therapy with low toxicity, reduced resistance, long-lasting viral suppression, less side effects, better oral bioavailability and improved potency[22, 23].

1.2 The HIV-1 virus

HIV is classified as a member of the *Lentiviris* subfamily of retroviruses[4, 8]. In 1983, HIV-1 was isolated by R.C. Gallo and L. Montagnier and few years later, HIV-2 was reported[1, 8, 24]. Both viruses cause AIDS but it was found that HIV-1 is the most widely spread virus worldwide.[8] The structure of the HIV-1 virus is 100 nm in diameter[8] (Figure 1). The viral particle is covered with a lipid-bilayer membrane traversed by surface glycoproteins (gp120), which are anchored via transmembrane proteins (gp41). The matrix shell protein (p17) forms the inside liner surface beneath the lipid-bilayer membrane. Capsid proteins (p24) form a cone-shaped arrangement that encloses the nucleocapsid proteins (p7 *gag* and p9 *gag*), reverse transcriptase, integrase and protease. HIV-1 carries three important genes namely, *gag, pol and env*[22]. These genes are articulated as large polyproteins which are processed into functional proteins by either the viral protease (*gag and pol*) or the host cellular enzymes (*env*). The *gag* gene encodes a precursor polyprotein that is cleaved to capsid proteins, matrix shell proteins and nucleocapsid proteins. The *pol* gene encodes for the three viral enzymes namely, reverse transcriptase, integrase and protease. The *env* gene encodes for the formation of the functional surface glycoproteins and transmembrane proteins[8].

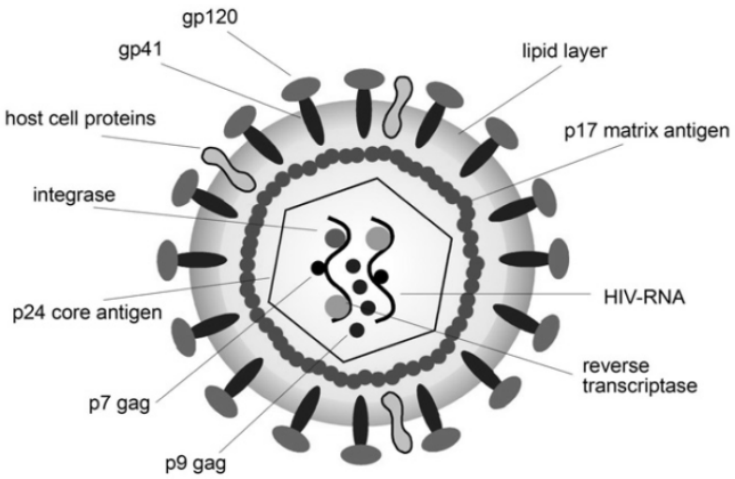


Figure 1. Structure of an HIV-1 virion particle[8].

1.2.1 HIV-1 life cycle

The HIV-1 virus infects T lymphocytes (T-cells) that carry a cluster of differentiation, namely the CD4 antigen, on their surface[4]. For the infection to be effective, it requires fusion with the viral and cellular membranes, see Figure 2[4]. When the virus enters the cell, its single-stranded ribonucleic acid (RNA) is reversed transcribed into double-stranded deoxyribonucleic acid (DNA) by reverse transcriptase in the host cytoplasm by means of the process called reverse transcription. This process is known to cause errors during the transcription process, resulting in frequent mutations for each generation of the virus produced.

Nucleoside reverse transcriptase inhibitors, nucleotide reverse transcriptase inhibitors and non-nucleoside reverse transcriptase inhibitors are designed to block the activity of the reverse transcriptase enzyme which are discussed in detail in Section 1.3.3[25].

The viral DNA is transported to the host nucleus, where the viral DNA is incorporated into the host DNA by the integrase enzyme forming a provirus. This process is called integration and is blocked by integrase inhibitors, which are discussed in Section 1.3.4. The virus replicates more frequently than the normal host DNA, without any detectable immune response and reaching concentrations of more than 100 million copies of the HIV-1 RNA/ml[8].

The activation of the host cell results in the transcription of the viral DNA into messenger RNA (mRNA) in the nucleus. The mRNA is transported to the host cytoplasm where it is translated into viral proteins. The products of translation are host cellular enzyme (env) precursor polyproteins (in the endoplasmic reticulum) and transmembrane proteins, *gag* and *gag-pol* protein precursors (in the ribosomes). All of these are then transported to the cell membrane. These proteins assemble in the membrane to form a curvature that results in the formation of a bud. HIV protease cleaves the new polyproteins, allowing them to join the RNA in the new functional viral proteins. The viral RNA

and the viral proteins gather at the surface into new virions where they bud from the host cell and are released to infect other host cells[4, 8, 26]. Therefore, protease inhibitors discussed in section 1.3.5 effectively prevent this stage due to the absence of sufficient and the correct viral proteins.

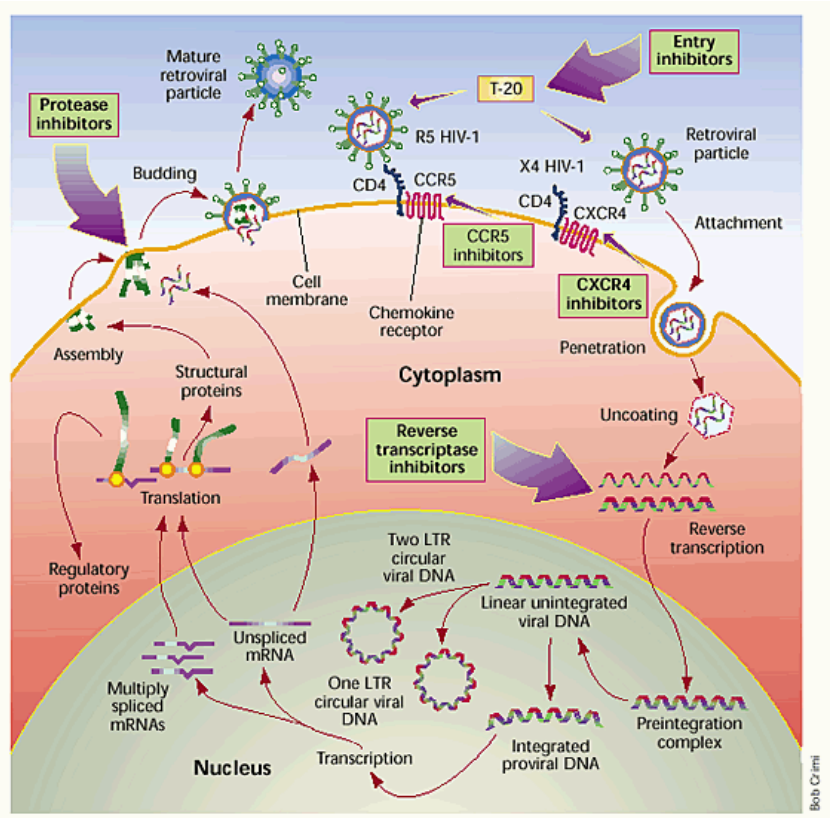


Figure 2. The replication cycle and targets for therapeutic intervention in the HIV life cycle[27].

Currently, there are several FDA-approved drugs commercially available for HIV/AIDS treatment.

1.3 Antiretroviral therapy

At present, a number of drugs are available on the market which is outlined in the following sections. The United States Food and Drug Administration approved the drugs on market and the year of drug approval is shown in brackets. These drugs are classified according to their therapeutic target in the HIV-1 replication cycle.

1.3.1 Types of enzyme inhibitors

Inhibitor types are classified based on the mechanism of action of inhibition. Inhibitors can also be broadly classified as specific and non-specific[28]. The first acts on a single enzyme only, while

the second type is effective against any enzyme. Non-specific inhibition can be caused by high temperature or extreme pH conditions and will lead to denaturing of the enzyme. It is, therefore, irreversible in nature.

Specific enzyme inhibition can be classified in reversible and irreversible inhibitors[28, 29].

The following are examples of reversible inhibitors.

- **Competitive inhibitors**

A competitive inhibitor is specific and has characteristics that diligently resemble the chemical and molecular geometry of the substrate thus, competing for the same enzyme active site as the substrate. The inhibitor binds precisely where the substrate binds thus, the bound inhibitor prevents the substrate from binding to the active site. This inhibition is reversible if sufficient substrate concentration is available to displace the competitive inhibitor from the active site[30, 31].

- **Non-competitive inhibitors**

A non-competitive inhibitor is also specific and binds to the free enzyme or enzyme-substrate but generally not at the active site. This leads to a change of the shape of the enzyme especially the active site so that the substrate can no longer effectively bind to the enzyme. Non-competitive inhibition is usually reversible and binds non-covalently to the enzyme[28, 30-32].

- **Uncompetitive inhibitors**

An uncompetitive inhibitor is structurally different from the substrate and it binds only to the enzyme-substrate complex. The binding of the inhibitor to the enzyme–substrate complex results in the loss of activity of the enzyme[30, 32].

- **Mixed inhibitors**

In this kind of inhibition, the inhibitor does not only compete with the substrate for binding to the enzyme but also bind to an enzyme molecule that subsequently binds a substrate or binds to an enzyme-substrate complex to affect catalytic turnover[32]. The binding of the inhibitor to the enzyme influences the binding of the substrate to the enzyme since it alters the protein's conformation[30].

Irreversible enzyme inhibition

Irreversible inhibitors form covalent interactions with the enzyme hence, forming a stable complex that is difficult to break. There are two types of irreversible enzyme inhibitors *i.e.*, the affinity

labels (active site-directed irreversible inhibitors) and the mechanism-based irreversible enzyme inactivators[28].

- **Affinity labels and active site-directed irreversible inhibitors**

The affinity labels are usually reactive chemical entities since they contain electrophilic groups. They target any nucleophilic group in the protein, especially those in the active site. These entities resemble the substrate structure and can bind to the enzyme thus, covalently modifying the enzyme. Normally, they tend to be unselective in their mode of action. Therefore, inherently toxic and they have little therapeutic value[28, 29].

- **Mechanism-based irreversible enzyme inhibitors**

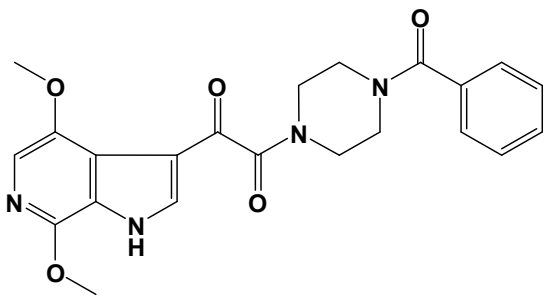
These kinds of inhibitors are also called suicide substrates. They are normally unreactive but become activated during catalytic processing. In their intermediate reactive state, they can irreversibly alkylate a nucleophilic amino acid group in the enzyme. This action causes the enzyme to denature (suicide)[28, 29].

1.3.2 Entry inhibitors

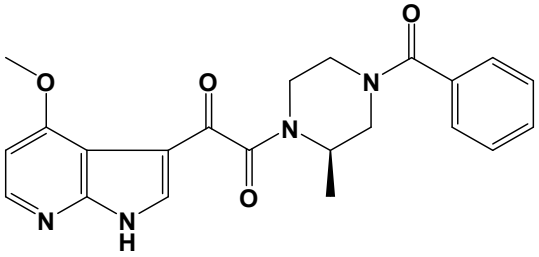
Entry inhibitors are competitive in character and prevent the entry of the virus into the host cell. In this section, entry inhibitors are explained according to their mode of action and examples are also shown.

1.3.2.1 Attachment or Binding inhibitors

These classes of compounds disrupt the initial binding or attachment of the HIV-1 gp120 surface protein to the CD4 T-cell surface receptor by one of the several possible mechanisms. Figure 3 shows some examples of clinically effective drugs[8].



BMS-488043



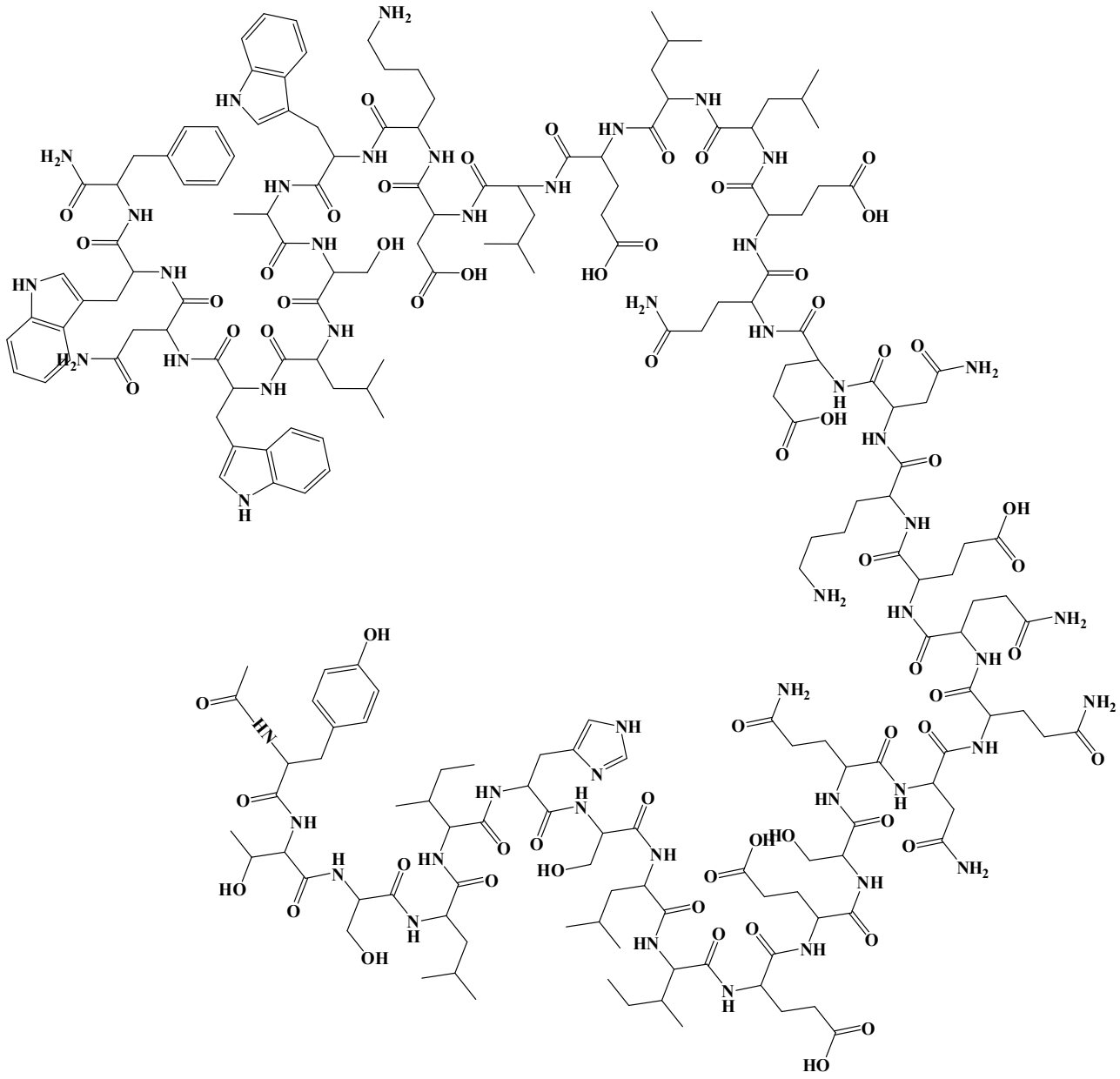
BMS-37880

Figure 3. Structures of attachment or binding HIV inhibitors[8, 33].

On the other hand, there are inhibitors designed to fuse with the host cell membrane. This kind of inhibitor is described in detail in the next section.

1.3.2.2 Fusion inhibitors

The first fusion inhibitor was approved by the FDA in 2003. It is a linear, 36-amino acid synthetic peptide which corresponds to residues 127-162 of the viral glycoprotein (gp41) named enfuvirtide[7, 34] (see Figure 4). This drug inhibits the conformational change of the gp41 which is required for fusion of the viral and host cell membrane. Therefore, the viral RNA is prevented from entering the host cell. Unfortunately, synthesis of the drug is difficult and expensive and it can only be administrated by subcutaneous injection[7].

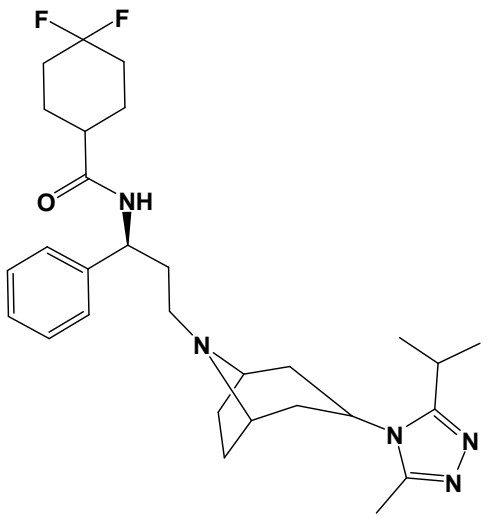


Enfuvirtide (2003)

Figure 4. The structure of a fusion inhibitor[6].

1.3.2.3 Co-receptor Antagonists inhibitors

There are two co-receptors namely, CCR5 and CXCR4 found on the surface of the T-cell. Co-receptor antagonists were the first antiretroviral drugs, which do not target the HIV virus directly, but they block the HIV attachment to the cell by preventing the virus from interacting with the co-receptors. Lead drugs targeting the CCR5 co-receptor include monoclonal antibody, Pro140 which blocks HIV-1 infection by occupying the gp120 binding site on CCR5[8]. Maraviroc was the first drug approved in 2007 that acts as a CCR5 antagonist (see Figure 5)[35].



Maraviroc (2003)

Figure 5. The structure of CCR5 antagonist maraviroc[35, 36].

HIV-1 reverse transcriptase enzyme is one of the important targets for antiviral therapy. The next section will describe the mechanism of action and associated FDA-approved drugs.

1.3.3 Reverse transcriptase inhibitors

These inhibitors interfere with the reverse transcription process of the viral RNA into DNA. In this section, reverse transcriptase inhibitors are explained according to their structural properties and mode of action and examples are also shown.

1.3.3.1 Nucleoside reverse transcriptase inhibitors

As mentioned earlier, AZT, a nucleoside reverse transcriptase inhibitor (NRTI) was the first drug approved in 1987 for treatment of HIV[7]. Unlike the natural deoxynucleotides substrates, NRTIs lack the 3'-hydroxyl group on the deoxyribose moiety[8]. These drugs are covalently bonded to the DNA. Incorporation of NRTI during DNA transcription prevents further covalent addition of nucleosides because the 5'-3' phosphodiester bond cannot be formed, therefore, transcription is terminated[8]. Currently, there are seven approved NRTIs on the market. Examples of some NRTIs are shown in Figure 6.

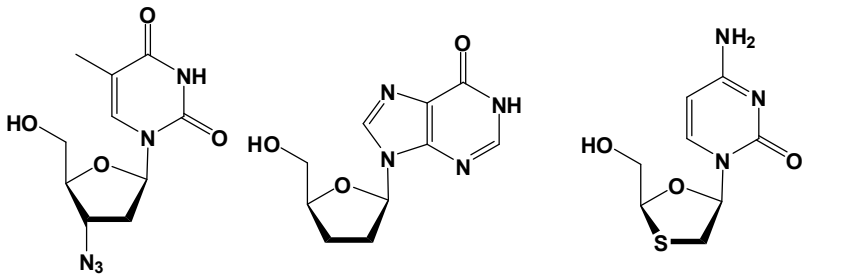


Figure 6. The structures of the nucleoside reverse transcriptase inhibitors approved by the FDA[7, 36].

The next section will outline other kinds of reverse transcriptase inhibitors, which exhibit less toxicity than nucleoside reverse transcriptase inhibitors.

1.3.3.2 Nucleotide reverse transcriptase inhibitors

The mode of action for nucleotide reverse transcriptase inhibitors is the same as with nucleoside reverse transcriptase inhibitors and they also form a covalent bond with DNA. The human body normally converts the nucleoside into a nucleotide through a phosphorylation process. The incorporation of a nucleotide skips the first phosphorylation process which might lead to toxicity[8]. Viread (Tenofovir), an example of a nucleotide analogue reverse transcriptase inhibitor, is shown in Figure 7.

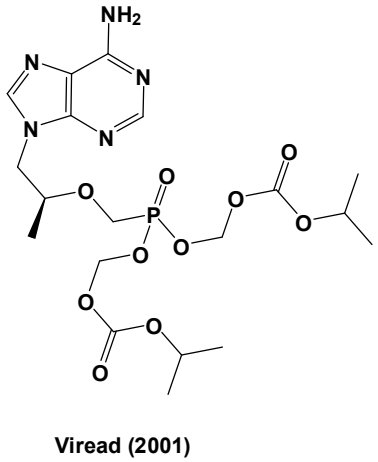


Figure 7. Structure of a nucleotide reverse transcriptase inhibitor[7, 36].

The next section describes another kind of reverse transcriptase inhibitor which is commercially available and form the part of HAART.

1.3.3.3 Non-nucleoside reverse transcriptase inhibitors

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) which were first described in 1990 interact specifically with the non-substrate binding site (allosteric site) on RT of HIV-1 is, therefore, it blocks its replication due to conformational changes at the active site of the enzyme[8]. Five NNRTIs were approved during the period of 1996-2011 (see Figure 8).

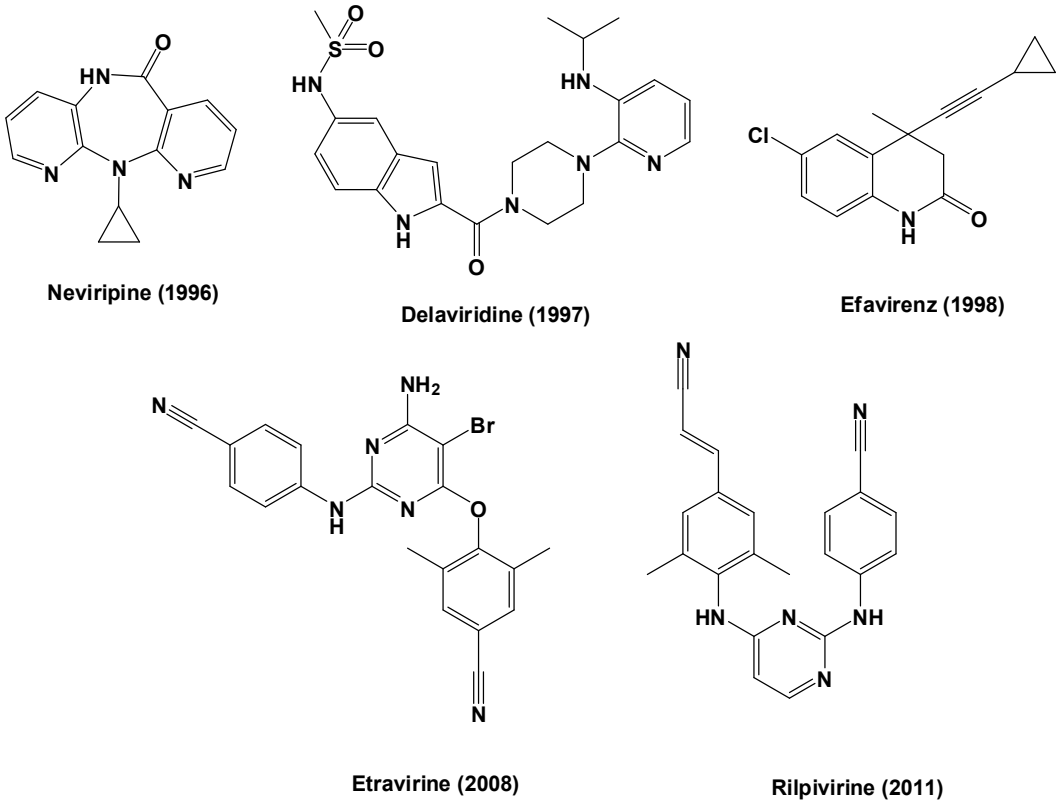


Figure 8. Structures of the non-nucleoside reverse transcription inhibitors[7, 35].

HIV-1 integrase enzyme is an important drug target and the next section will describe the mode of action of integrase inhibitors.

1.3.4 Integrase inhibitors

The development and advancement of the research on HIV-1 integrase inhibitors has lagged behind for almost a decade because of insufficient knowledge of the integrase structure. Integrase inhibitors block the integrase enzyme, which is essential for the HIV-1 replication cycle by blocking the integration of the viral DNA into the host genome. Raltegravir was the first integrase inhibitor to be approved by FDA in 2007 (see Figure 9)[8, 37, 38].

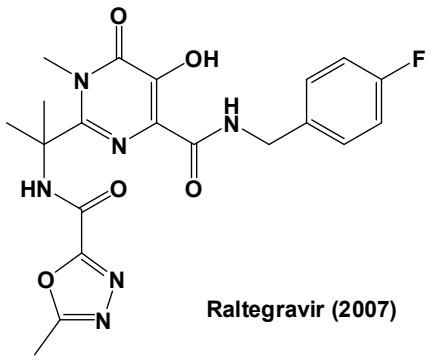


Figure 9. Structure of integrase inhibitor approved by FDA[8, 37-39].

The HIV-1 protease also plays an important role in the HIV-1 life cycle since it is responsible for viral maturation. The next section describes the mode of action of the PR inhibitors.

1.3.5 Protease inhibitors

The clinical practice of HIV protease inhibitors (PIs) (see Figure 10) was brought into use in 1995[23]. In 1995, saquinavir was the first PR drug approved for clinical practice[23]. The advantages of HIV-1 protease inhibitors include prolonged viral control, better viral suppression, reduction of the indication of HIV infection and reduction of death rates[23]. HIV PR treatment is, therefore, very beneficial to patients infected with HIV. A sound understanding of the virus and its life cycle have attracted researchers' interest in the discovery of new PIs[8]. The mechanism of action for the drug is to inhibit the catalytic activity of the protease then preventing new virions from maturing. This renders the virions into a non-infectious state[8]. One limitation with HIV PIs is resistance which leads to disease proliferation and ineffective treatment[23].

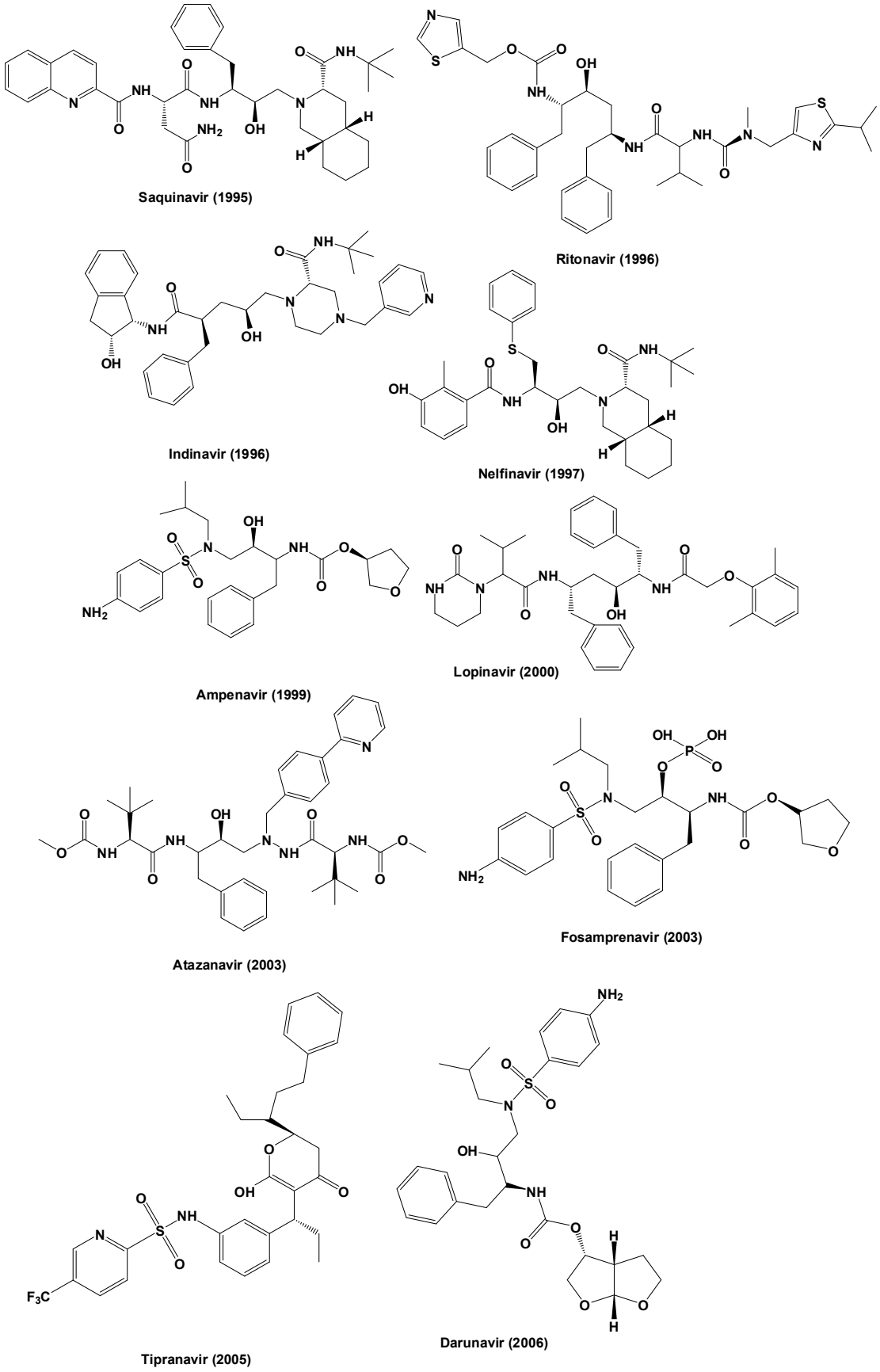


Figure 10. Structures of the HIV-1 protease inhibitors approved by the FDA[35, 39].

A comparative study of the binding affinities was recently reported for nine FDA-approved drugs (ritonavir, amprenavir, saquinavir, indinavir, nelfinavir, lopinavir, darunavir, tipranavir and atazanavir) against HIV-1 subtype B PR and C-SA HIV-1 PR. It was found that from the nine investigated drugs, most inhibitors exhibited slightly weaker binding affinities for C-SA HIV-1 PR than subtype B PR[40]. Therefore, development of new drugs with better binding affinities is needed for the C-SA HIV-1 PR.

In this study, pentacycloundecane-based compounds as HIV-1 PR inhibitors will be looked at in more detail.

Other developments in terms of HIV drug research

A recent review article outlined the use of nanopharmaceuticals as potential anti-HIV therapies[26]. Nanopharmaceuticals such as nanocrystals, nanocapsules, nanoparticles, solid lipid nanoparticles, nanocarriers, micelles, liposomes and dendrimers have been studied. Some of these nanopharmaceuticals exhibit antiviral activity by targeting different aspects of the HIV-1 life cycle. However there are limitations to the delivery of drugs with nanoparticles due to biological challenges, toxicity, cost and difficulties for scaling up[26]. Nanopharmaceuticals is a developing field of research.

1.4 HIV-1 protease enzyme

The fully functional form of the HIV-1 protease that is needed for the virus to survive is the homodimeric form (22 kDa) and not the monomeric (11 kDa), inactive form. The protease cleaves newly synthesized polyproteins to produce mature HIV virion[4].

1.4.1 The structure of HIV-1 protease

The HIV-1 protease is classified as a C_2 symmetric active homodymeric aspartyl protease which is essential for the HIV-1 life-cycle[14, 23]. The enzyme consists of 99 amino acids per monomer and the position of each monomer forms an axis of symmetry[14]. The secondary structure of each monomer consists of one α-helix and two antiparallel β sheets. There are three domains in the enzyme shown in Figure 11 which is referred to as the active site cavity (the core domain), terminal domain (dimerization domain) and the flap domain[41]. From generation to generation, it has been observed that the nucleotide sequence of HIV-1 protease changes[4]. However, the tripeptide sequence found in each monomer *i.e.* Asp25-Thr26-Gly27 is conserved[14]. The enzyme is characterized by its ability to cleave peptidic substrates with the help of Asp25 and Asp25' residues

found in each subunit. Therefore, the two Asp25 residues act as the catalytic residues during hydrolysis[4, 8, 14].

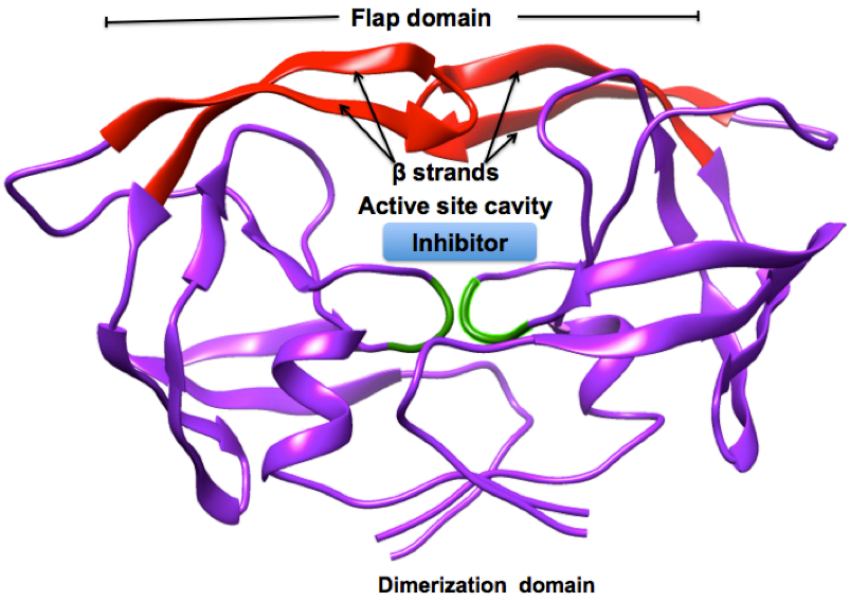


Figure 11. Structure of HIV-1 protease.

The HIV-1 PR has three domains which are described in the next section.

- **Active site cavity (core domain)**

  The tripeptide Asp25-Thr26-Gly27 is found at the interface of the core domain formed by the two monomers. This domain plays an important role in the stabilization of the dimer and the catalytic site[4, 5, 14].
- **Terminal domain (dimerization domain)**

  The terminal domain consists of antiparallel β sheets which are formed by the four termini in the dimer of residues 1-4 and 95-99[34]. This domain is important in dimer formation and stabilization of an active protease.
- **The flap domain**

  Flaps are two hairpin turn structures which are made up of amino acid residues 43-58 found in both monomers[14, 42-44]. These flexible regions cover the active site and have a great effect on ligand/inhibitor binding interactions. It has been revealed from the crystal structure of the HIV-1 protease that the flaps exhibit great mobility ranging from closed form, semi-open form and wide-open form[42, 43, 45]. In the closed form state the active site is occupied by the ligand while the wide-open form state allows the active site to be accessed by the substrate and the inhibitor[42, 43, 45]. This section is described in detail in Chapter 3.

Although the HIV-1 PR has been crystallized as a single monomer (accession 1HHP pdb code),[46] it functions in the dimeric form. The dimer is observed in complexes of the HIV-PR with inhibitors. The dimer is formed due to intermolecular forces that contribute to its stability and the next section will describe these forces.

1.4.2 The stability of HIV-1 protease

The aliphatic residues found in the PR dimer stabilize each monomer in a hydrophobic core[34, 47]. In addition, the stabilization of the HIV protease dimer is achieved by factors such as non-covalent interactions, hydrophobic packing of side chains and interaction involving the catalytic residues[47]. A hydrogen bonding network present in the catalytic site is shown in Figure 12[5, 14].

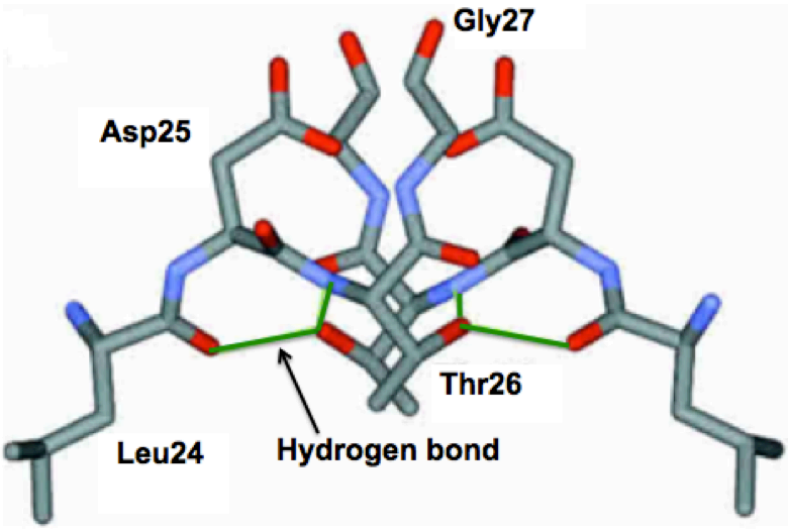


Figure 12. The catalytic residues of the tripeptide of Asp25-Thr26-Gly27 forming a hydrogen bond network (green lines) [47].

1.4.3 Substrate binding

The active site pocket is made up with several well-defined subsites where side chains of the substrate can be accommodated upon binding. The subsites for each monomer are named S_1 to S_n and S_1' to S_n' (the prime symbol, $_1$ to $_n$ denotes the direction from the N-terminal to the C-terminal direction[5]) while the amino acid residues of the substrate are named as P_1 to P_n and P_1' to P_n', when counting from scissile bond (see Figure 13) that is cleaved during hydrolysis[48]. As a result of the C_2 symmetric character of the protease subsites S_1 and S_1' are the same and this holds for the other subsites[4, 49, 50].

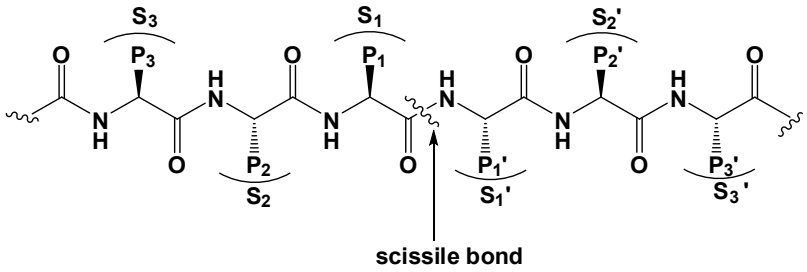


Figure 13. The standard nomenclature representing P_1 _ _ _ P*n*, P_1'_ _ _ _ P*n*'_ for amino acid residues of peptide substrates. The corresponding binding sites on the protease are referred to as S_1_ _ _ S*n*, S_1'_ _ _ _ S*n*'_ subsites[4].

1.4.4 HIV-1 protease catalytic mechanism

Proteases are degradative enzymes, which catalyze the hydrolysis of peptide bonds with high sequence selectivity and catalytic proficiency[4, 48]. There are two broad classes of protease enzymes, i.e., exopeptidases and endopeptidases. These proteases are classified according to their catalytic mechanism[4]. The exopeptidases hydrolyze the peptide bond nearer to the amino or carboxyl termini of the substrate, while endopeptidases hydrolyze the peptide bonds far away from the termini of the substrate[48]. These two classes of enzymes are further classified into four groups, i.e., serine proteases, aspartic proteases, cysteine proteases and metalloproteases, based on the functional groups present at the catalytic site[48]. HIV-1 protease is an example of aspartic endopeptidase proteases since it has Asp25-Thr26-Gly27 residues at the catalytic site. HIV-1 protease cleaves the *gag* and *gag-pol* protein by an acid-base mechanism (see Figure 14)[13, 14, 51]. The aspartic protease catalyzes the hydrolysis of the peptide bond by activation of a water molecules which then attacks the amide bond carbonyl of the substrate's scissile bond. This results in an oxy-anion tetrahedral intermediate (see Figure 14)[4, 14]. Activation of one of the water molecule results due to the combined action of the two aspartyl β-carboxyl groups at the active site[4, 25]. The protonation of the amide nitrogen is followed by a breakdown of the tetrahedral intermediate resulting in the formation of a carboxylic acid and an amine as the final products[4, 52].

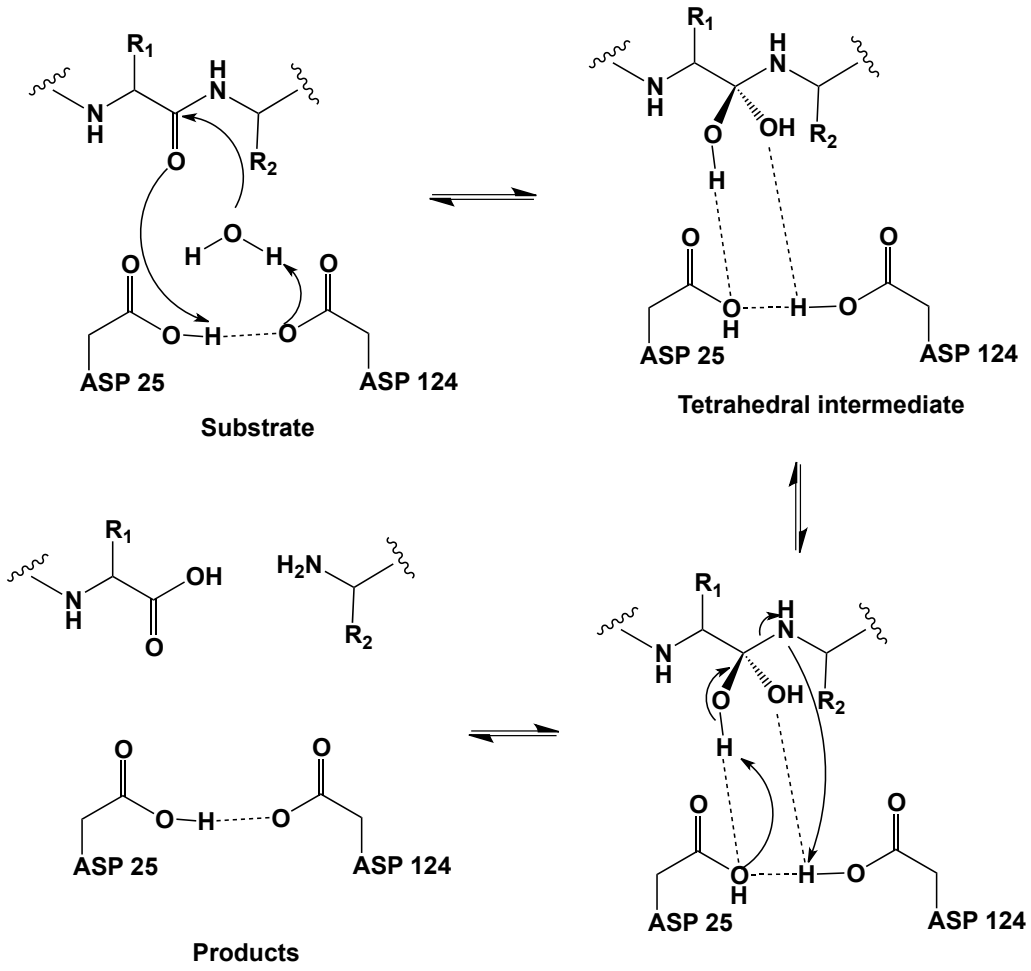


Figure 14. Schematic representation of the cleavage mechanism of HIV-1 protease[4, 25].

Scope of this study

Polycyclic cage compounds have been reviewed intensively[53-57] and have shown antiviral, pharmacokinetic and pharmacodynamic properties which are explained in more detail in Section 1.5.

In this study, we employed molecular dynamics (MD) simulation studies using Amber software, to investigate the subtype C South African (C-SA) HIV-1 protease motion and also, Autodock 4.0 was used for docking the PCU-based compounds in the active site of the C-SA HIV-1 protease. A background for computational studies is presented in Chapter 2.

1.5 Polycyclic cage compounds

Polycyclic cage molecules have been extensively studied as potential chemical scaffolds for the development of new drugs due to the broad range antiviral properties amantadine had shown[58, 59]. Incorporation of the cage into bioactive molecules has revealed pharmacokinetic and pharmacodynamic properties such as the improvement of drug transportation across membranes due

to the hydrophobicity of the hydrocarbon cage. In addition, the steric effect and rigidity of the cage results in the metabolic degradation process being retarded thus, prolonging the activity and reducing the frequency of drug administration to the patient. The rigidity of the cage also induces receptor site specificity in antibacterial activity, anabolic action and analgesic activity[25, 58]. The introduction of advantageous properties by cage molecules has inspired our group to develop new potent polycyclic cage peptides for potential treatment of HIV-1.

Peptides show potent pharmacological properties and are specifically recognized by protease enzymes[52]. In our group, protease inhibitors with the pentacycloundecane (PCU) frame work with either peptides or peptoids as side chains have been synthesized, and tested for activity[60-70]. The first study involved the synthesis and biological testing of PCU-lactam peptides against C-SA HIV-1 PR and the inhibitor reported an IC_{50} of 78 nM[60, 61]. The second study involved a PCU-ether peptide with the inhibitor exhibiting an IC_{50} = 600 nM[66]. In the third study, PCU-lactam peptoids were reported with similar activities (500 nM). The fourth study involved the synthesis of PCU-diol peptoids and the inhibitor with IC_{50} = 7500 nM was reported[64]. The fifth study involved the synthesis PCU-diol peptides and the inhibitor with IC_{50} = 500 nM was reported[63].

In this study, PCU-based compounds were investigated by employing computational methods such as molecular docking (see Section 2.7) and molecular dynamic simulations (see Section 2.9.)

1.6 References

1. F. Barré-Sinoussi, J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axlerblin, F. Vezinetbrun, C. Rouzioux, W. Rozenbaum, L. Montagnier. Isolation Of A T-Lymphotropic Retrovirus From A Patient At Risk For Acquired Immune-Deficiency Syndrome (AIDS). *Science* 1983,**220**:868-871.
2. W. Blattner, R. C. Gallo, H. M. Temin. HIV Causes AIDS. *Science* 1988,**241**:515-515.
3. V. Hornak, C. Simmerling. Targeting Structural Flexibility In HIV-1 Protease Inhibitor Binding. *Drug Discovery Today* 2007,**12**:132-138.
4. A. Brik, C. Wong. HIV-1 Protease : Mechanism And Drug Discovery. *Organic and Biomolecular Chemistry* 2003,**1**:5-15.
5. A. Wlodawer, J. Vondrasek. Inhibitors Of HIV-1 Protease: A Major Success Of Structure-Assisted Drug Design. *Annual Review of Biophysics and Biomolecular Structure* 1998,**27**:249-286.
6. In: http://www.avert.org/aids-histroy-86.htm 04 April 2011.
7. R. J. Pomerantz, D. L. Horn. Twenty Years Of Therapy For HIV-1 Infection. *Nature Medicine* 2003,**9**:867-875.
8. C. Hoffmann, J. K. Rockstroh. HIV 2010 *Medizin Fokus Verlag* 2010.
9. Joint United Nations Programme On HIV/AIDS World AIDS Day Report 2011. In: http://www.unaids.org 14 November 2012.
10. H. Mitsuya, K. J. Weinhold, P. A. Furman, M. H. St. Clair, S. N. Lehrman, R. C. Gallo, D. Bolognesi, D. W. Barry, S. Broder. 3'-Azido-3'-deoxythymidine (BW A509U): An Antiviral Agent That Inhibits The Infectivity And Cytopathic Effect Of Human T-Lymphotropic Virus Type III/Lymphadenopathy-Associated Virus In Vitro. *Proceedings of the National Academy of Sciences of the United States of America* 1985,**82**:7096-7101.
11. M. Barry, F. Mulcahy, C. Merry, S. Gibbons, D. Back. Pharmacokinetics And Potential Interactions Amongst Antiretroviral Agents Used To Treat Patients With HIV Infection. *Clinical Pharmacokinetics* 1999,**36**:289-304.
12. A. R. Zolopa. The Evolution Of HIV Treatment Guidelines: Current State-Of-The-Art Of Art. *Antiviral Research* 2010,**85**:241-244.

13. J. M. Lenhard, D. K. Croom, J. E. Weiel, D. A. Winegar. HIV Protease Inhibitors Stimulate Hepatic Triglyceride Synthesis. *Arteriosclerosis Thrombosis and Vascular Biology* 2000,**20**:2625-2631.

14. H. C. Castro, P. A. Abreu, R. B. Geraldo, R. C. A. Martins, R. D. Santos, N. I. V. Loureiro, L. M. Cabral, C. R. Rodrigues. Looking At The Proteases From A Simple Perspective. *Journal of Molecular Recognition* 2011,**24**:165-181.

15. J. Fellay, K. Boubaker, B. Ledergerber, E. Bernasconi, H. Furrer, M. Battegay, B. Hirschel, P. Vernazza, P. Francioli, G. Greub, M. Flepp, A. Telenti. Prevalence Of Adverse Events Associated With Potent Antiretroviral Treatment: Swiss HIV Cohort Study. *Lancet* 2001,**358**:1322-1327.

16. R. G. Jain, E. S. Furfine, L. Pedneault, A. J. White, J. M. Lenhard. Metabolic Complications Associated With Antiretroviral Therapy. *Antiviral Research* 2001,**51**:151-177.

17. A. Carr. Toxicity Of Antiretroviral Therapy And Implications For Drug Development. *Nature Reviews Drug Discovery* 2003,**2**:624-634.

18. G. Chene, J. A. Sterne, M. May, D. Costagliola, B. Ledergerber, A. N. Phillips, F. Dabis, J. Lundgren, A. D. Monforte, F. de Wolf, R. Hogg, P. Reiss, A. Justice, C. A. Deminie, S. Staszewski, J. Gill, G. Fatkenheuer, M. E. Egger, A. R. T. C. Collaboration. Prognostic Importance Of Initial Response In HIV-1 Infected Patients Starting Potent Antiretroviral Therapy: Analysis Of Prospective Studies. *Lancet* 2003,**362**:679-687.

19. M. A. Martinez. Progress in the Therapeutic Applications of siRNAs Against HIV-1. In: *Methods in Molecular Biology*. Edited by Sioud, M.; 2009. pp. 343-368.

20. B. A. I. Payne, I. J. Wilson, C. A. Hateley, R. Horvath, M. Santibanez-Koref, D. C. Samuels, D. A. Price, P. F. Chinnery. Mitochondrial Aging Is Accelerated By Anti-Retroviral Therapy Through The Clonal Expansion Of mtDNA Mutations. *Nature Genetics* 2011,**43**:806-811.

21. D. J. Kempf, K. C. Marsh, J. F. Denissen, E. Mcdonald, S. Vasavanonda, C. A. Flentge, B. E. Green, L. Fino, C. H. Parks, X. Kong, N. E. Wideburg, A. Saldivar, L. Ruizll, W. M. Kati, H. L. Sham, T. Robins, K. D. Stewarti, A. Hsu, J. J. Plattner, J. M. Leonardtt, D. W. Norbeck. ABT-538 Is A Potent Inhibitor Of Human Immunodeficiency Virus Protease And Has High Oral Bioavailability In Humans. *Proceedings of the National Academy of Sciences of the United States of America* 1995,**92**:2488-2493.

22. J. P. Vacca, J. H. Condra. Clinically Effective HIV-1 Protease Inhibitors. *Drug Discovery Today* 1997,**2**:261-273.

23. J. L. Martinez-Cajas, M. A. Wainberg. Protease Inhibitor Resistance In HIV-infected Patients: Molecular And Clinical Perspectives. *Antiviral Research* 2007,**76**:203-222.

24. R. C. Gallo, P. S. Sarin, E. P. Gelmann, M. Robertguroff, E. Richardson, V. S. Kalyanaraman, D. Mann, G. D. Sidhu, R. E. Stahl, S. Zollapazner, J. Leibowitch, M. Popovic. Isolation Of Human T-Cell Leukemia-Virus In Acquired Immune-Deficiency Syndrome (AIDS). *Science* 1983,**220**:865-867.
25. M. M. Makatini. Design, Synthesis And Screening Of Novel PCU-Peptide/Peptiod Derived HIV Protease Inhibitors [PhD thesis]. Durban: University of KwaZulu-Natal; 2011.
26. R. Parboosing, Maguire, G. E. M., P. Govender, Kruger, H. G. Nanotechnology And The Treatment Of HIV Infection. *Viruses-Basel* 2012,**4**:488-520.
27. N. L. Michael, J. P. Moore. HIV-1 Entry Inhibitors : Evanding The Tissue. *Nature Medicine* 1999,**5**:740-742.
28. J. M. Berg, J. L. Tymoczko, Stryer, L. *Biochemistry*. 5th ed: W H Freeman; 2002.
29. T. L. Lemke, D. A. Williams, V. F. Roche, W. S. Zito. *Foye's Principles Of Medicinal Chemistry*. 7th ed; 2012.
30. R. H. Garrett, C. M. Grisham. *Biochemistry*. 4th ed: Mary Finch; 2010.
31. M. Roberts, M. Reiss, G. Monger. *Advanced Biology*: Nelson; 2000.
32. D. Zhang, M. Zhu, G. W. Humphreys. *Drug Metabolism In Drug Design And Development*: John Wiley & Sons, Inc; 2008.
33. H. Ho, L. Fan, B. Nowicka-Sans, B. McAuliffe, C. Li, G. Yamanaka, N. Zhou, H. Fang, I. Dicker, R. Dalterio, Y. Gong, T. Wang, Z. Yin, Y. Ueda, J. Matiskella, J. Kadow, P. Clapham, J. Robinson, R. Colonno, P. Lin. Envelope Conformational Changes Induced By Human Immunodeficiency Virus Type 1 Attachment Inhibitors Prevent Cd4 Binding And Downstream Entry Events. *Journal of Virology* 2006,**80**:4017-4026.
34. V. Briz, E. Poveda, V. Soriano. HIV Entry Inhibitors: Mechanisms Of Action And Resistance Pathways. *Journal of Antimicrobial Chemotherapy* 2006,**57**:619-627.
35. FDA. Antiretroviral Drugs Used In The Treatment Of HIV Infection. In: http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm; 27 June 2013.
36. De Clercq, E. Antiretroviral Drugs. *Current Opinion in Pharmacology* 2010,**10**:507-515.
37. Y. Pommier, A. A. Johnson, C. Marchand. Integrase Inhibitors To Treat HIV/AIDS. *Nature Reviews Drug Discovery* 2005,**4**:236-248.
38. B. Dau, M. Holodniy. Novel Targets For Antiretroviral Therapy *Drugs* 2009,**69**:31-52.
39. B. T. Nguyen, R. D. Isaacs, H. Teppler, R. Y. Leavitt, P. Sklar, M. Iwamoto, L. A. Wenning, M. D. Miller, J. Chen, R. Kemp, W. Xu, R. A. Fromtling, J. P. Vacca, S. D. Young, M. Rowley, M. W. Lower, K. M. Gottesdiener, D. J. Hazuda. Raltegravir: The First HIV-1 Integrase Strand Transfer Inhibitor In The HIV Armamentarium. In: *Pharmaceutical*

Science to Improve the Human Condition: Winners and Finalist Candidates of the Prix Galien USA, International, and Pro Bono Humanitarian Awards 2010. Edited by Braaten, D., Cooney, R. E.; 2011. pp. 83-89.

40. S. M. Ahmed, H. G. Kruger, T. Govender, G. E. M. Maguire, Y. Sayed, M. Ibrahim, P. Naicker, M. E. S. Soliman. Comparison Of The Molecular Dynamics And Calculated Binding Free Energies For Nine FDA-approved HIV-1 PR Drugs Against Subtype B and CSA HIV PR. *Chemical Biology and Drug Design* 2013,**81**:208-218.
41. I. T. Weber, J. Agniswamy. HIV-1 Protease: Structural Perspectives On Drug Resistance. *Viruses* 2009,**1**:1110-1136.
42. W. R. P. Scott, C. A. Schiffer. Curling Of Flap Tips In HIV-1 Protease As A Mechanism For Substrate Entry And Tolerance Of Drug Resistance. *Structure* 2000,**8**:1259-1265.
43. V. Hornak, A. Okur, R. C. Rizzo, C. Simmerling. HIV-1 Protease Flaps Spontaneously Close To The Correct Structure In Simulations Following Manual Placement Of An Inhibitor Into The Open State. *Journal of the American Chemical Society* 2006,**128**:2812-2813.
44. V. Hornak, A. Okur, R. C. Rizzo, C. Simmerling. HIV-1 Protease Flaps Spontaneously Open And Reclose In Molecular Dynamics Simulations. *Proceedings of the National Academy of Sciences of the United States of America* 2006,**103**:915-920.
45. G. Verkhivker, G. Tiana, C. Camilloni, D. Provasi, R. A. Broglia. Atomistic Simulation Of The HIV-1 Protease Folding Inhibition. *Biophysical Journal* 2008,**95**:550-562.
46. S. Spinelli, Q. Z. Liu, P. M. Alzari, P. H. Hirel, R. J. Poljak. The 3-Dimensional Structure Of The Aspartyl Protease From The HIV-1 Isolate BRU. *Biochimie* 1991,**73**:1391-1396.
47. M. Ingr, T. Uhlikova, K. Strisovsky, E. Majerova, J. Konvalinka. Kinetics Of The Dimerization Of Retroviral Proteases: The "Fireman's Grip" And Dimerization. *Protein Science* 2003,**12**:2173-2193.
48. M. B. Rao, A. M. Tanksale, M. S. Ghatge, V. V. Deshpande. Molecular And Biotechnological Aspects Of Microbial Proteases. *Microbiology and Molecular Biology Reviews* 1998,**62**:597-635.
49. J. Tozser. Comparative Studies On Retroviral Proteases: Substrate Specificity. *Viruses* 2010,**2**:147-165.
50. I. Schechter, A. Berger. On The Size Of The Active Site In Proteases. *Biochemical and Biophysical Research Communications* 2012,**425**:497-502.
51. J. Ekegren. Design And Synthesis Of Novel HIV-1 Protease Inhibitors Comprisinga Tertiary Alcohol In The Transition-State Mimic [PhD thesis]: Uppsala University; 2006.

52. J. Nguyen, Y. Hamada, T. Kimura, Y. Kiso. Design Of Potent Aspartic Protease Inhibitors To Treat Various Diseases. *Archiv der Pharmazie* 2008,**341**:523-535.

53. D. W. Oliver, S. F. Malan. Medicinal Chemistry Of Polycyclic Cage Compounds In Drug Discovery Research. *Medicinal Chemistry Research* 2008,**17**:137-151.

54. A. P. Marchand. Polycyclic Cage Compounds: Reagents, Substrates, And Materials For The 21St Century. *Aldrichimica Acta* 1995,**28**:95-104.

55. Suri, S. C. Polynitro Polycyclic Cage Compounds. In: University of Dayton; 1993. pp. 267-285.

56. J. Joubert, W. J. Geldenhuys, C. J. Van der Schyf, D. W. Oliver, H. G. Kruger, T. Govender, S. F. Malan. Polycyclic Cage Structures As Lipophilic Scaffolds For Neuroactive Drugs. *Chemical Medicinal Chemistry* 2012,**7**:375-384.

57. W. J. Geldenhuys, S. F. Malan, J. R. Bloomquist, A. P. Marchand, C. J. Van der Schyf. Pharmacology And Structure-Activity Relationships Of Bioactive Polycyclic Cage Compounds: A Focus On Pentacycloundecane Derivatives. *Medicinal Research Reviews* 2005,**25**:21-49.

58. K. Bisetty. A Computational Study Of Cage Peptides [PhD thesis]. Durban: University of KwaZulu-Natal; 2002.

59. W. J. Geldenhuys, S. F. Malan, J. R. Bloomquistd, C. J. Van der Schyf. Structure–Activity Relationships Of Pentacycloundecylamines At The N-Methyl-D-Aspartate Receptor. *Bioorganic & Medicinal Chemistry* 2007,**15**:1525-1533.

60. M. M. Makatini, K. Petzold, S. N. Sriharsha, M. E. S. Soliman, B. Honarparvar, P. I. Arvidsson, Y. Sayed, P. Govender, G. E. M. Maguire , H. G. Kruger, T. Govender. Pentacycloundecane-Based Inhibitors Of Wild-Type C-South African HIV-Protease. *Bioorganic & Medicinal Chemistry Letters* 2011,**21**:2274-2277.

61. M. M. Makatini, K. Petzold , S. N. Sriharsha , N. Ndlovu , M. E.S. Soliman , B. Honarparvar , R. Parboosing , A. Naidoo , P. I. Arvidsson, Y. Sayed , P. Govender , G. E.M. Maguire , H. G. Kruger , T. Govender. 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* 2011,**46**:3976-3986.

62. B. Honarparvar, H. G. Kruger, G. E. M. Maguire, T. Govender, M. E. S. Soliman. 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 Reviews* 2013 Accepted.

63. B. Honarparvar, M. M. Makatini, K. Petzold, M. E. S. Soliman, P. I. Arvidsson, Y. Sayed, T. Govender, G. E. M. Maguire, H. G. Kruger. Pentacycloundecane-Diol Based HIV-1 Protease Inhibitors: Biological Screening, 2D-NMR And Molecular Simulations Studies. *Chemical Medicinal Chemistry* 2012,**7**:1009-1020.

64. M. Makatini, K. Petzold, P. I. Arvidsson, B. Honarparvar, T. Govender, G. E. M. Maguire, R. Parboosing, Y. Sayed, M. E. S. Soliman, H. G. Kruger. Synthesis, Screening And Computational Investigation Of Pentacycloundecane-Peptoids As Potent CSA-HIV PR Inhibitors. *European Journal of Medicinal Chemistry* 2012 **57**:459-468.

65. M. Makatini, K. Petzold, Alves., C. N., N. Ndlovu, B. Honarparvar, P. Govender, T. Govender, H. G. Kruger, Y. Sayed, J. Lameira, G. E. M. Maguire, Soliman, M. E. S. Synthesis, 2D-NMR And Molecular Modelling Studies Of Pentacyclo-Undecane Lactam Peptides And Peptoids As Potential HIV-1 Wild type C-SA Protease Inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry* 2013,**28**:78-88.

66. R. Karpoormath, Y. Sayed, P. Govender, T. Govender, H. G. Kruger, M. E. S. Soliman, G. E. M. Maguire. 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* 2012,**40**:19-30.

67. R. Karpoormath, F. Albericio, T. Govender, G. E. M. Maguire, H. G. Kruger. Synthesis And NMR Elucidation Of Pentacycloundecane-Derived 3 Hydroxy Acid Peptides As Potential Anti-HIV-1 Agents. *Structural Chemistry* 2013 Accepted.

68. R. Karpoormath, Y. Sayed, T. Govender, H. G. Kruger, M. E. S. Soliman, G. E. M. Maguire. Novel PCU Cage Diol Peptides As Potential Targets Against Wild-Type CSA HIV-1 Protease: Synthesis, Biological Screening And Molecular Modelling Studies. *Medicinal Chemistry Research* 2013 Accepted.

69. O. K. Onajole, T. Govender, M. Makatinia, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Penta-Cycloundecane Amine Derivatives As Potential Antituberculosis Agents. *Magnetic Resonance in Chemistry* 2008,**46**:1007-1014.

70. O. K. Onajole, M. M. Makatini, P. Govender, T. Govender, G. E. M. Maguire, Kruger, H. G. Synthesis And NMR Assignment Of Pentacycloundecane Precursors Of Potential Pharmaceutical Agents. *Magnetic Resonance in Chemistry* 2010,**48**:249-255.

Chapter 2

Introduction to computational chemistry

Theoretical methods used in this thesis

This chapter outlines the computational and theoretical chemistry tools used in this study. A recent review from our group describes the combination of drug design, molecular modelling and spectroscopic techniques[1]. This review also presents a thorough overview of available computational techniques for this purpose.

There are two basic methods of computational chemistry; namely, quantum mechanics and molecular dynamics. The first is based on electronic structure methods, solving the Schrödinger equation and the second on the classical laws of physics. The first is much more costly in terms computational resources and time and is therefore, only practical for smaller systems (up to 100s of atoms), while the second can handle much larger biochemical systems of tens of thousands of atoms[2]. Hybrid method such QM/MM/MD will also be discussed in Section 2.10.

2.1 Quantum Mechanics

In everyday life, basic principles of classical mechanics is applied but fails when applied to the transfer of very small energies and to objects with very small masses; therefore, quantum mechanics originated[3-5]. In quantum mechanics, all properties of a system are expressed in terms of a wave function that is obtained by solving the Schrödinger equation[3].

2.2 Schrödinger equation

In 1926 Erwin Schrödinger proposed an equation referred to as the time-independent Schrödinger equation for a particle of mass *m* moving in one dimension with energy, *E*, which is given as[3, 6-8],

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi = E\psi \qquad \text{Eq. 1}$$

Where *V(x)* is associated with the potential energy of the particle at point *x*; *m* is the particle mass; ψ is the wave function of the quantum system; *E* is the sum of potential and kinetic energies and *ħ* is the Planck's constant, 6.626 x 10^{-34}Js.

The Schrödinger time-independent wave equation can also be defined by combining the classical differential equation which describes the profile of a simple harmonic standing wave and De' Broglie's relation between matter and waves[6, 8, 9].

$$\left[-\frac{h^2}{8\pi^2}\sum_i \frac{1}{m_j}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V(r)\right].\psi(r)=E.\psi(r) \qquad \text{Eq. 2}$$

Where $\psi(r)$ is related with any set of i particles with masses m_j and $V(r)$ is the potential field in which the particles are moving and r refers to the set of component vectors of each particle as a function of x, y and z coordinates. **Eq.2** can also be written in its abbreviated form:

$$H\psi = E\psi \qquad \text{Eq.3}$$

In the Schrödinger equation, H is the Hamiltonian operator which is the sum of the kinetic energy *(T)* and the potential energy *(V)* operators[7-9].

$$H = T + V \qquad \text{Eq. 4}$$

Where H is defined as follows:

$$H = \left[-\frac{h^2}{8\pi^2}\sum_i \frac{1}{m_j}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\right]+\sum_i\sum_{<j}\left(\frac{e_i e_j}{r_{ij}}\right) \qquad \text{Eq. 5}$$

The Born-Oppenheimer approximation (see Section 2.3) was then introduced, when using Eq. 5 could not solve molecular systems.

2.3 Born-Oppenheimer approximation

The Born-Oppenheimer approximation shows how the electronic motion can be approximately separated from the nuclear motion[6, 7]. It tries to separate the electronic and nuclear degrees of freedom even though they are coupled by the electron-nuclear potential energy. In most cases, the Hamiltonian operator can be simplified by using the Born-Oppenheimer approximation, since the mass of the nucleus is much greater than that of an electron and the nuclei move very slowly with respect to the electron while the electrons react quickly to changes in nuclear positions[7-9]. Thus, the electron distribution within a molecular system can be approximately described if the nuclei have fixed positions while their kinetic contributions in the kinetic operator are not taken into consideration[8, 9].

$$T^{elec} = \left[-\frac{h^2}{8\pi^2 m}\sum_i^{electrons}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\right] \qquad \text{Eq. 6}$$

Whereby the Schrödinger equation for electrons in the fixed nuclei is written as:

$$H^{elec}\psi^{elec}(r,R) = E^{eff}(R)\psi^{elec}(r,R) \qquad \text{Eq. 7}$$

Where $E^{eff}(R)$ is the effective electronic energy which depends on the relative nuclear coordinates R. The solutions of electronic Hamiltonian H^{elec} are the electronic wave function $\psi^{elec}(r,R)$ where r

is the electronic coordinates and R is the nuclear coordinates. Solving this equation for other fixed positions of interest will produce a Potential Energy Surface (PES)[6, 9]. Specific points of interest on the PES are minima (low energy conformations) and maxima (saddle points or transition states).

2.4 Potential Energy Surface

A PES is a mathematical function that requires the systematic calculation of energies of a molecule or system as a function of its geometrical arrangements[2, 3]. Quantum mechanical computational techniques take into account electronic correlation which arises from instantaneous interactions between electrons as they move closer and further from each other in a molecular orbital[3]. A model surface of the energy as a function of the molecular geometry and interesting features are outlined in Figure 15. The most interesting points on the PES's are the stationary points where the gradients with respect to all internal coordinates are zero. Stationary point includes local or global minima which corresponds to the low energy equilibrium molecular structures and saddle points[2]. Saddle points are the lowest energy barriers on the path connecting minima that relates to the concept of a transition state[2].

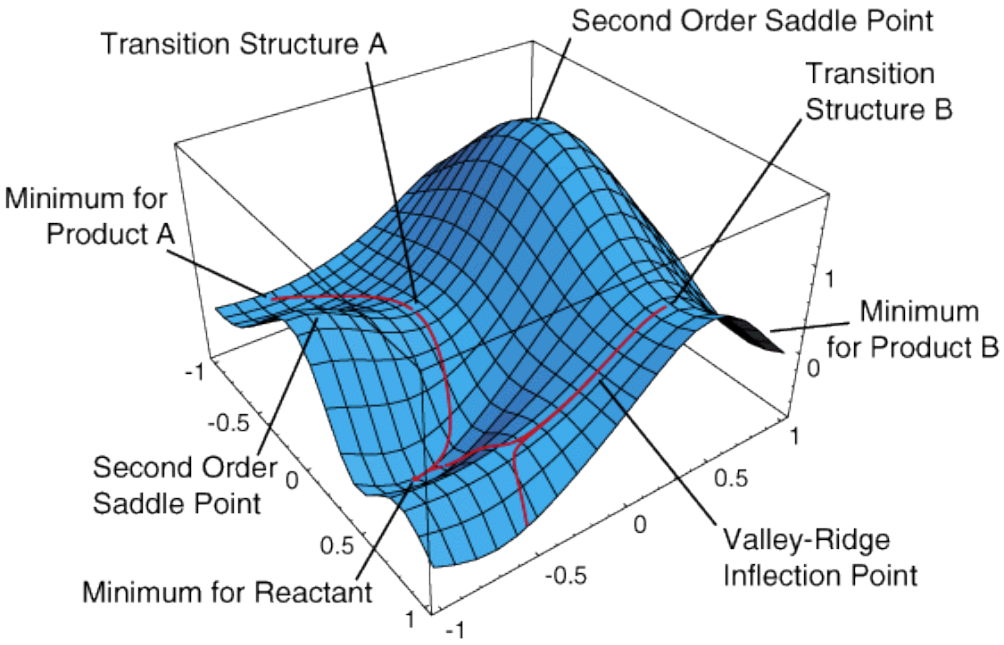


Figure 15. An example of two-dimensional potential energy surface[10].

2.5 Molecular Mechanics

Molecular mechanics (MM) uses the law of classical physics to explain and interpret the behaviour or properties of atoms and molecules from a single geometry[3, 5, 7, 11]. The advantage of MM

over quantum mechanics is that it relies on force-fields with embedded empirical parameters and can be used for molecules as large as enzymes because it does not calculate electronic properties[5, 7]. Conformational analysis or energy minimization is perhaps the single most important application of molecular mechanics[12]. The main components of a force field comprise of a set of equations explaining how the potential energy of the molecule varies with locations of its component atoms, and a series of atom types defined by the characteristics of an atom of the element within a specific chemical context[11].

There are drawbacks that accompany molecular mechanics models[5, 7]. First, they are limited to the description of equilibrium geometries and equilibrium conformations. Second, it does not calculate electronic properties; therefore, it gives limited information about the bonding. It cannot be used for the calculation of bond formation/breaking. Thus, chemical reactivity and selectivity cannot be predicted. Third, the force fields used mostly nowadays are parameterized using experimental data or alternative data from *ab initio* techniques. Last, the accuracy of molecular mechanics depends not only on the choice of parameters but also on experimental studies and the parameterization, which is normally based on small number of model systems[5, 7, 12]. It is important to note that a specific force field is designed for a specific family of molecules (proteins etc.) and it may not work for a completely different system.

MM consider a molecule to be a collection of masses interacting with each other through harmonic forces. Thus, the MM total energy of a molecule is defined by the sum of contribution resulting from bonded interactions such as bond length, bond angles, torsions and non-bonded interactions such as Van der Waals and electrostatic[6, 8, 11, 12].

$$\boldsymbol{E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec}} \ldots. \qquad \text{Eq. 8}$$

Where E_{tot} is the total energy, E_{str} is the bond-stretching energy, E_{bend} is the angle-bending energy, E_{tors} is the torsional energy, E_{vdw} is the Van der Waals energy and E_{elec} is the electrostatic energy. The combination of such sets of energies constitutes molecular mechanics force fields[7-9, 13, 14].

2.5.1 Force fields

Numerous molecular mechanics force fields have been developed and used today. Force fields such as AMBER, CHARMM, OPLS and GROMOS are the most popular and the most used set of parameters applied for simulation for biomolecules[12, 15]. All force fields have different parameters due to differences of the parameterization process, the atom types and the reference data. Thus, these force fields are referred to as set of variables that must be adjusted to give results of a similar global good fitting[12, 15].

In this study, AMBER[16] force fields were used. The General AMBER Force Field (GAFF) is used to parameterize the ligands while the standard AMBER force field was used for the protease.

2.6 Homology models

Homology modelling is also known as comparative modelling of proteins such as enzymes[17]. The lack of high resolution, three-dimensional (3D) structural information of proteins from X-ray crystallography, multidimensional NMR spectroscopy and other experimental methods resulted into the development of homology modelling. This method refers to the construction of a target protein from its amino acid sequence and 3D structural information of a related homologous protein used as a template[1, 17]. There are several steps normally required for constructing a model[1].

- Search and select:

A known 3D structure with about 25% or better similarity in the primary structure than the target protein with unknown 3D is searched and selected

- Align:

The sequences of the known and unknown proteins are aligned. During alignment the structurally preserved and variable regions will be identified.

- Model:

The target protein coordinates for the structurally preserved regions are modelled from those of the protein template with known 3D structure. The conformation of the loops and side chains are also constructed without the template and this might results in inaccuracy.

- Evaluate:

The constructed model is then evaluated.

In this study, homology modelling was not used to model C-SA HIV-1 PR from subtype B HIV-1 PR. This was not required since the required changes were manually introduced. The construction of C-SA HIV-1 PR is described in detail in Chapter 3. The crystal structure of C-SA HIV-1 was recently solved[18] and gave an excellent correlation with the model we have created.

Molecular docking method is used for the prediction of the bound geometry and interactions of small organic molecules (ligands, substrates) with their biological targets (receptors such as enzymes)[19]. Normally, it is recommended to employ virtual screening before experimental work to save time and costs. The next section will describe molecular docking.

2.7 Molecular Docking

Computational methods are becoming more important in studying the structure and function of biomolecules due to the number of protein X-ray structures or disease targets known today[20-22].

Molecular docking is mostly used in structure-based rational drug design to identify correct conformations of small molecule ligands and also to estimate the strength of the protein-ligand interaction. Molecules such as substrates, inhibitors or other drug candidates are identified in the active site of the macromolecule. Macromolecules are proteins such as a receptor, enzyme or nucleic acid with a known 3-dimensional structure[23, 24]. Molecular docking forms a link between the structural component and the synthetic compound in rational drug design[25]. The binding energy between the ligand and the macromolecule is analyzed by using molecular mechanics[20]. Molecular mechanics is explained in detail in Section 2.5

The binding energy is calculated as follows:

$$\boldsymbol{E_{binding} = E_{target} + E_{ligand} - E_{target-ligand}} \qquad \text{Eq. 9}$$

There are more than 60 different docking programs for both commercial and academic work [23] such as Dock[26], AutoDock[27], FlexX[28], Surflex[29], Gold[30], Glide[31], Cdocker[32], LigandFit[33], MCDock[34] and others. The proposed docking programs can be categorized into four broad categories namely, stochastic Monte Carlo, fragment-based, evolutionary-based and the shape complementary methods[23]. Docking programs do not use a systematic search to explore all degrees of freedom fully because of the computational costs[23].

There are two types of *docking* namely, rigid and flexible docking. In rigid docking both macromolecule and ligand are kept rigid while in flexible docking the flexibility is allowed for the macromolecule (or selected part thereof) or the ligand or both[35]. In this study, we have employed AutoDock 4.0[36] for docking a flexible ligand in the active site of the rigid HIV-1 protease to estimate the strength of the protein-ligand interaction. AutoDock is an evolutionary method that uses the genetic algorithm to perform conformational searches and it uses the Lamarckian version of genetic algorithm. The Lamarckian genetic algorithm [27] can handle ligands with larger number of degrees of freedom[37]. In the Lamarckian genetic algorithm new conformations are adapted from changes in the conformation of molecules after *in situ* optimization and the strength of the interaction between the ligand and the protein would be determined[23].

Mukherjee *et al.,* employed three different docking experiments, i.e., rigid docking, fixed anchored docking and flexible docking[19]. From the three experiments, it was reported that maximum errors were from scoring function at 1.5-2.0 Å. Therefore, it was concluded that general failures in docking are significantly due to scoring functions than sampling[19]. Until now, there is no scoring method developed to consistently identify the correct binding mode for any given protein-ligand complex[5]. Scoring failures in docking measures the inability of the energy function to allocate the best score to a correctly sampled conformation out of the ensemble generated. A more accurate energy function might improve the overall results. It was also revealed that the number of rotatable

bonds affects the results. The success rate for ligands containing rotatable bonds of 7-0, 8-15 and 15 plus, can be ranked as 7-0> 8-15> 15 plus[19]. In this study, PCU-diol peptides with ≥ 12 rotatable bonds were investigated. Thus, these compounds have a high probability of rendering poor results. As the ligand flexibility increases its success rate during docking decreases[19]. The use of a single approach such as a hybrid method implemented by combining different existing docking techniques might maximizes advantages and minimizes disadvantages[38].

2.8 The Austin Model 1 (AM1) semi-empirical approximation

In semi-empirical methods, many of the integrals are estimated by using quantum physics, or experimentally-derived empirical parameters. It also uses a series of rules to set certain integrals equal to zero[3, 4, 7, 8]. In this study, AM1 BCC charges[39] are used during ligand preparation. The AM1 method is based on the Neglect of Diatomic Differential Overlap (NDDO)[2, 12, 40]. If two electrons are defined by two basic functions then the two-electron integral is:

$$\int\int \emptyset_a(\mathbf{1})\emptyset_b(\mathbf{1})\left(\frac{\mathbf{1}}{r_{12}}\right)\emptyset_c(\mathbf{2})\emptyset_d(\mathbf{2})d\tau_1 d\tau_2 \qquad \text{Eq. 10}$$

Where basic functions a and b are used to define electron 1 and basic functions c and d are used to define electron 2[7]. Examples are shown in Figure 16, Figure 17 and Figure 18.

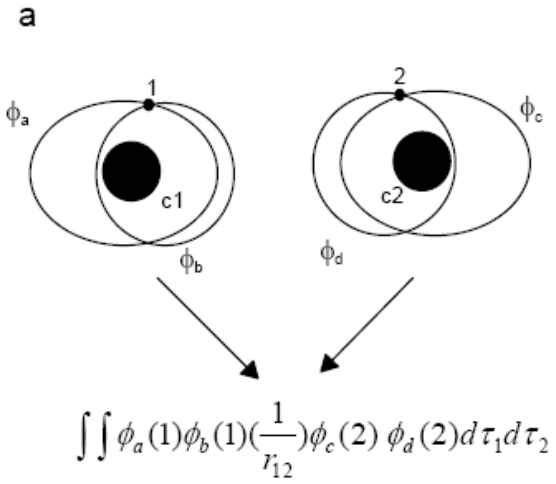


Figure 16. The AM1 evaluation for two-center two electron interaction integrals where electron repulsion integrals are calculated[12].

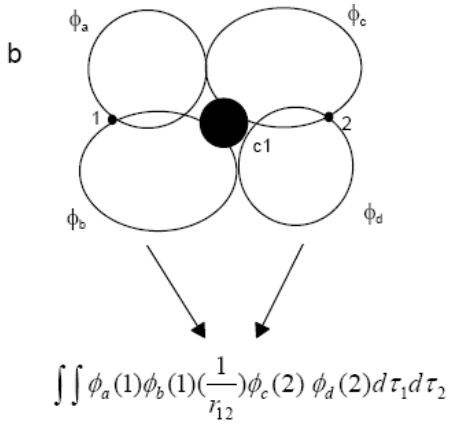


Figure 17. The AM1 evaluation for one-center two electron interaction integrals[12].

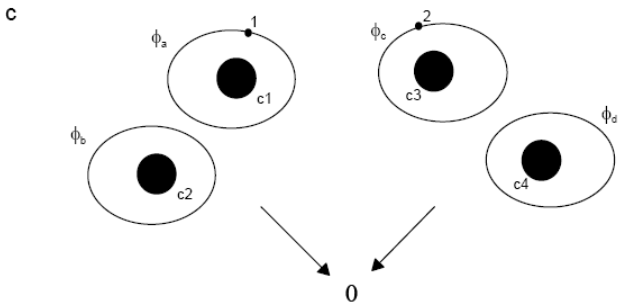


Figure 18. A case where AM1 method is not applicable because it omits two-electron repulsions over three and four centers[12].

After assigning atomic charges and other system settings with AMBER, a MD simulation is performed. The following section will explain how the MD method works.

2.9 Molecular Dynamics

A molecular dynamics (MD) simulation is defined as a form of computer simulation programme which aims to reproduce the time-dependent behaviour[9, 41, 42] of a molecule due to atom vibrations, movements and collisions that results from the interactions between the atoms and molecules for a period of time by approximations of known physical attributes for a system of particles at a given temperature[4, 11]. In a molecular dynamic simulations the molecule is set in motion by heating to a specific temperature to overcome potential energy barriers and to achieve more conformations at high temperature[3, 14]. Molecular dynamics calculations are based on molecular mechanics principles[14, 43, 44]. The conformations are generated by integrating Newton's laws of motion[8, 9, 45]. The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time[4, 9, 11]. This is achieved by finding the

force on each particle (F_i) first as a function of time, which is equal to the negative gradient of the potential energy:

$$\boldsymbol{F_i} = -\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{r_i}} \qquad \text{Eq. 11}$$

Where *U* is the potential energy function and *r* is the position of the particle. From Newton's Law of motion, acceleration *(a)* of a particle is determined by dividing the force acting on it by the mass of the particle[4, 8].

$$\boldsymbol{a_i} = -\frac{\boldsymbol{F_i}}{\boldsymbol{m_i}} \qquad \text{Eq. 12}$$

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

$$\boldsymbol{dv} = \int \boldsymbol{adt}, \qquad \text{Eq. 13}$$

$$\boldsymbol{dr} = \int \boldsymbol{vdt} \qquad \text{Eq. 14}$$

Lastly the kinetic energy can be described in terms of both the velocities and momenta of the particles[4, 8].

$$\boldsymbol{K(v)} = \frac{\boldsymbol{1}}{\boldsymbol{2}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{m_i v_i^2}, \qquad \text{Eq. 15}$$

$$\boldsymbol{K(p)} = \frac{\boldsymbol{1}}{\boldsymbol{2}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \frac{\boldsymbol{p_i^2}}{\boldsymbol{m_i}} \qquad \text{Eq. 16}$$

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

$$\boldsymbol{H(q,p)} = \boldsymbol{K(p)} + \boldsymbol{U(q)} \qquad \text{Eq. 17}$$

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

$$\boldsymbol{v_i(t)} = \frac{\boldsymbol{d}}{\boldsymbol{dt}} \boldsymbol{q_i(t)} \qquad \text{Eq. 18}$$

Where $q_i(t)$ refers to the atomic positions at a particular time, *t*. Based on the initial atom coordinates of the system, new positions and velocities of the atoms at time *t* and the atoms will move to these new positions. Thus, new conformations are generated[8, 11]. The temperature of the system is directly proportional to the average kinetic energy.

In order to obtain more accurate and precise results hybrid methods are employed. The next section will explain QM/MM/MD as one of the hybrid methods.

2.10 Hybrid method (QM/MM/MD)

In this methodology, the system is divided into two parts. The small reactive part of the chemical system is defined by quantum mechanics (QM) whereas the remaining large non-reactive part is defined by molecular mechanics (MM) and treated with a force-field calculation (see Figure 19)[5]. A MD simulation is then applied to the entire system. The advantage of this methodology is that it gives valuable information in qualitative analysis while it is not good for quantitative analysis[1, 5, 12, 46].

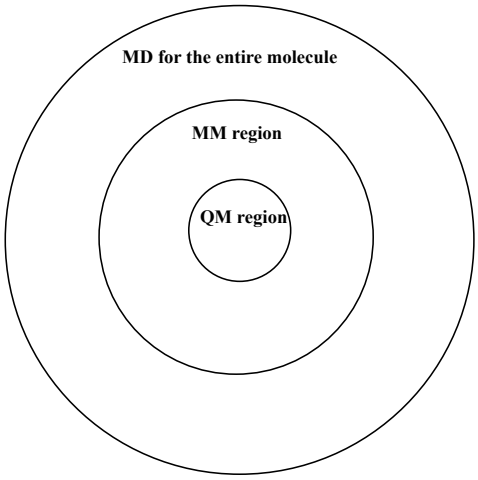


Figure 19. An example of a hybrid QM/MM/MD model.

Factors that affect quantitative analysis when using this methodology[5]**:**

- Requires good *ab initio* description for the QM part
- Requires extensive sampling while accessing free binding energy numbers which is computationally expensive.
- Require calibration of the interaction between the QM part and the MM part which is difficult for biochemical systems such as enzymes.

With the results from MD simulation, the binding free energy of complexes can be calculated. The next section will describe how the binding free energy is calculated.

2.11 Binding free energy

The binding free energy of a molecular system plays an important role in predicting the ability to associate and react[44, 47]. In this study, the binding free energy calculations are performed for the ligands binding to C-SA HIV-1 PR enzyme by using the Molecular mechanics Generalized Born Surface Area (MM-GBSA) method implemented in Amber 11[16].

The binding free energy ΔG_b is calculated as follows:

$$\boldsymbol{\Delta G_b = \Delta G_{MM} + \Delta G_{sol} - T\Delta S} \quad \text{Eq. 19}$$

Where ΔG_{MM} is the molecular mechanics (MM) representing the interaction energy between the protein and the ligand, ΔG_{sol} is the solvation free energy and $T\Delta S$ is for the entropy contribution to the binding[48, 49].

The ΔG_{MM} term is calculated from the MM interaction energies as follows:

$$\boldsymbol{\Delta G_{MM} = \Delta G_{elecint} + \Delta G_{vdwint}} \quad \text{Eq. 20}$$

Here, $\Delta G_{elecint}$ and ΔG_{vdwint} are the electrostatic and Van der Waals interaction energies between the protein and the ligand, respectively[48, 49].

On the other hand the ΔG_{sol} term can be separated into two parts as follows:

$$\boldsymbol{\Delta G_{sol} = \Delta G_{elecsol} + \Delta G_{nonpolarsol}} \quad \text{Eq. 21}$$

In this case, $\Delta G_{elec\ sol}$ and $\Delta G_{\ nonpolar\ sol}$ are the electrostatic and nonpolar contribution of the solvation free energy and nonpolar contributions to the solvation free energy is calculated as a linear function of the solvent accessible surface area (SA)[48, 49].

$$\boldsymbol{\Delta_{nonpolarsol} = \gamma SA + b} \quad \text{Eq. 22}$$

where $\gamma = 0.00542$ kcal mol^{-1}Å^{-2} and $b = 0.92$ kcal/mol.

2.12 References

1. B. Honarparvar, H. G. Kruger, G. E. M. Maguire, T. Govender, M. E. S. Soliman. 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 Reviews* 2013 Accepted.
2. C. J. Cramer. *Essentials Of Computational Chemistry: theories and models* 2nd ed: John Wiley and Sons, Ltd; 1961.
3. P. Atkins, J. de Paula. *Atkins' Physical Chemistry*: Oxford University Press; 2006.
4. K. B. Lipkowitz, D. B. Boyd. *Reviews In Computational Chemistry*: VCH Publishers, Inc; 1990.
5. P. Bultinck, H. de Winter, W. Langenaeker, J. P. Tollenaere. *Computational Medicinal Chemistry For Drug Discovery*: Marcel Dekker, Inc; 2004.
6. A. Hinchliffe. *Molecular Modeling For Beginners*: John Wiley and sons Ltd; 2003.
7. D. W. Rogers. *Computational Chemistry Using PC*. 3rd ed: John Wiley & Sons, Inc; 2003.
8. F. Jensen. *Introduction To Computational Chemistry*: John wiley & Sons; 1999.
9. K. Bisetty. A Computational Study Of Cage Peptides [PhD thesis]. Durban: University of KwaZulu-Natal; 2002.
10. In: http://www.chem.wayne.edu/~hbs/chm6440/PES.html; 21 February 2013.
11. P. Sharma. An Assessment Of The Conformational Profile Of Bombensin And Its Mammalian Analogues Using Computational Chemistry Methods [PhD thesis]. Durban: Durban University of Technology; 2011.
12. M. E. S. Soliman. Computational Modelling Of Glycosidase Mechanisms: Structural And Mechanistic Aspects [PhD thesis]. United Kingdom: University of Bath; 2011.
13. S. J. Weiner, P. A. Kollman, D. A. Case, C. U. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. Weiner, P. Weiner. A New Force-Field For Molecular Mechanical Simulation Of Nucleic-Acids And Proteins. *Journal of the American Chemical Society* 1984,**106**:765-784.
14. C. E. Sansom, C. A. Smith. Computer Applications In The Biomolecular Sciences. Part1: Molecular Modelling. *Biochemical Education* 1998,**26**:103-111.
15. G. Keserü, I. Kolossváry. Molecular Mechanics And Conformational Analysis In Drug Design. In. 1 ed: Blackwell Science Ltd.; 1999.
16. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, 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, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, Kollman, P. A. AMBER 11. In. San Francisco: University of California; 2010.

17. M. A. Johnson, B. M. Pinto. NMR Spectroscopic And Molecular Modeling Studies Of Protein-Carbohydrate And Protein-Peptide Interactions. *Carbohydrate Research* 2004,**339**:907-928.
18. P. Naicker, I. Achilonu, S. Fannuchi, M. Fernandes, M. A. A. Ibrahim, H. W. Dirr, Soliman, M. E. S., Y. Sayed. Structural Insights Into The South African HIV-1 Subtype C Protease: Importance Of Hinge Region Dynamics And Flap Flexibility In Drug Resistance. *Journal of Biomolecular Structure and Dynamics* 2012 Manuscript accepted.
19. S. Mukherjee, T. E. Balius, R. C. Rizzo. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *Journal of Chemical Information and Modeling* 2010,**50**:1986-2000.
20. A. Rudnitskaya, B. Török, M. Török. Molecular Docking Of Enzyme Inhibitors. *Biochemistry And Molecular Biology Education* 2010,**34**:261-266.
21. S. Kortagere, M. D. Krasowski, S. Ekins. The Importance Of Discerning Shape In Molecular Pharmacology. *Trends in Pharmacological Sciences* 2009,**30**:138-147.
22. T. P. Lybrand. Ligand Protein Docking And Rational Drug Design. *Current Opinion in Structural Biology* 1995,**5**:224-228.
23. D. Plewczynski, M. Lazniewski, R. Augustyniak, K. Ginalski. Can We Trust Docking Results? Evaluation Of Seven Commonly Used Programs On PDBbind Database. *Journal of Computational Chemistry* 2010,**32**:742-756.
24. M. Botta, F. Corelli, F. Manetti, A. Tafi. Molecular Modeling As A Powerful Technique For Understanding Small-Large Molecules Interactions. *Farmaco* 2002,**57**:153-165.
25. A. J. Olson, D. S. Goodsell. Automated Docking And The Search For HIV Protease Inhibitors. *SAR and QSAR in Environmental Research* 1998,**8**:273-285.
26. T. J. A. Ewing, S. Makino, A. G. Skillman, I. D. Kuntz. Dock 4.0: Search Strategies For Automated Molecular Docking Of Flexible Molecule Databases. *Journal of Computer-Aided Molecular Design* 2001,**15**:411-429.
27. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson. Automated Docking Using A Lamarckian Genetic Algorithm And An Empirical Binding Free Energy Function. *Journal of Computational Chemistry* 1998,**19**:1639-1663.
28. M. Rarey, B. Kramer, T. Lengauer, G. Klebe. A Fast Flexible Docking Method Using An Incremental Construction Algorithm. *Journal of Molecular Biology* 1996,**261**:470-490.

29. A. N. Jain. Surflex: Fully Automatic Flexible Molecular Docking Using A Molecular Similarity-Based Search Engine. *Journal of Medicinal Chemistry* 2003,**46**:499-512.

30. G. Jones, P. Willet, R. C. Glen, A. R. Leach, R. Taylor. Development And Validation Of A Genetic Algorithm For Flexible Docking. *Journal of Molecular Biology* 1997,**267**:727-749.

31. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin. Glide: A New Approach For Rapid, Accurate Docking And Scoring. 1. Method And Assessment Of Docking Accuracy. *Journal of Medicinal Chemistry* 2004,**47**:1739-1750.

32. G. Wu, D. H. Robertson, C. L. Brooks, M. Vieth. Detailed Analysis Of Grid-Based Molecular Docking: A Case Study Of CDOCKER—A CHARMm-Based MD Docking Algorithm. *Journal of Computational Chemistry* 2003,**24**:1549-1562.

33. C. M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman. Ligandfit: A Novel Method For The Shape-Directed Rapid Docking Of Ligands To Protein Active Sites. *Journal of Molecular Graphics and Modelling* 2003,**21**:289-308.

34. M. Liu, S. Wang. MCDOCK: A Monte Carlo Simulation Approach To The Molecular Docking Problem. *Journal of Computer-Aided Molecular Design* 1999,**13**:435-452.

35. J. A. Erickson, M. Jalaie, D. H. Robertson, R. A. Lewis, M. Vieth. Lessons In Molecular Recogntion : The Effect Of Ligands And Protein Flexibility On The Molecular Docking Accuracy. *Journal of Medicinal Chemistry* 2004,**47**:45-55.

36. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson. AutoDock4 And AutoDockTools4: Automated Docking With Selective Receptor Flexibility. *Journal of Computational Chemistry* 2009,**30**:2785-2791.

37. M. L. Teodoro, G. N. Phillips Jr, L. E. Kavraki. Molecular Docking : A Problem With Thousands Of Degrees Of Freedom. 2001,**1**:960-966.

38. J. C. Mitchell, A. T. Phillips, J. B. Rosen, L. F. Ten Eyck. A Coupled Scanning And Optimization Scheme For Analyzing Molecular Interactions. In: *Optimization in Computational Chemistry and Molecular Biology*. Edited by C. A. Floudas, P. M. Pardalos: Kluwer Academic Publishers B.V.; 2000. pp. 190-207.

39. A. Jakalian, D. B. Jack, C. I. Bayly. Fast, Efficient Generation Of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization And Validation. *Journal of Computational Chemistry* 2002,**23**:1623-1641.

40. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart. AM1 - A New General-Purpose Quantum-Mechanical Molecular-Model. *Journal of the American Chemical Society* 1985,**107**:3902-3910.

41. M. T. H. Khan. Molecular Interactions Of Cholinesterases Inhibitors Using In Silico Methods: Current Status And Future Prospects. *New Biotechnology* 2009,**25**:331-346.

42. W. Wang, O. Donini, C. M. Reyes, P. A. Kollman. 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* 2001,**30**:211-243.

43. P. Cozzini, M. Fornabaio, A. Marabotti, D. J. Abraham, G. E. Kellogg, A. Mozzarelli. Free Energy Of Ligand Binding To Protein: Evaluation Of The Contribution Of Water Molecules By Computational Methods. *Current Medicinal Chemistry* 2004,**11**:3093-3118.

44. S. C. L. Kamerlin, M. Haranczyk, A. Warshel. 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* 2009,**113**:1253-1272.

45. R. D. Taylor, P. J. Jewsbury, J. W. Essex. A Review Of Protein-Small Molecule Docking Methods. *Journal of Computer-Aided Molecular Design* 2002,**16**:151-166.

46. P. Amara, M. J. Field, C. Alhambra, J. Gao. The Generalized Hybrid Orbital Method For Combined Quantum Mechanical/Molecular Mechanical Calculations: Formulation And Tests Of The Analytical Derivatives. *Theoretical Chemistry Accounts* 2000,**104**:336-344.

47. P. Kollman. Free-Energy Calculations - Applications To Chemical And Biochemical Phenomena. *Chemical Reviews* 1993,**93**:2395-2417.

48. H. Meiselbach, A. H. C. Horn, T. Harrer, H. Sticht. Insights Into Amprenavir Resistance In E35D HIV-1 Protease Mutation From Molecular Dynamics And Binding Free-Energy Calculations. *Journal of Molecular Modeling* 2007,**13**:297-304.

49. G. Hu, Q. Zhang, L. Y. Chen. Insights Into scFv: Drug Binding Using The Molecular Dynamics Simulation And Free Energy Calculation. *Journal of Molecular Modeling* 2011,**17**:1919-1926.

Chapter 3

Comprehensive molecular dynamics flap movement analysis of the HIV PR free enzyme and complex with the natural substrate and various inhibitors

3.1 Abstract

HIV protease (PR) flap flexibility is crucial for substrate entry to access the catalytic site for effective hydrolysis of the natural substrate. In this study, we have proposed and tested several metrics for obtaining optimum analysis of the flap dynamics. Molecular dynamic simulations have been employed to reach four goals. First, comparison of the flap flexibility between subtype B and C-SA HIV-1 PR was carried out to provide greater insight into the impact of polymorphisms on the enzyme flexibility. Second, different approaches to estimate the extent of flap dynamics of PR have been adapted to accurately measure flap flexibility. Third, the impact of different classes of PCU-based inhibitors on the flap dynamics of C-SA HIV PR was studied and compared to both the natural substrate as well as ritonavir. Last, to investigate to what extent the natural flap dynamics should reflect critical qualitative information about the nature of the interaction between the inhibitor and the HIV PR. Results were compared against the measured IC_{50} for the synthesized PCU-based inhibitors. It was observed that an implicit solvent MD model explores the complete conformational profile in a much shorter time (10 ns) in comparison to explicit solvent systems. The metric parameter for measuring the distance between the Cα K55 – Cα K154 gave the best measure for monitoring flap movement since it is not affected by flap-curling and flap asymmetry. The overall results showed that C-SA PR has higher flexibility in the flaps compared to subtype B strain both for the free enzyme and inhibitor/enzyme complexes. When looking at the MD results and the flap movement measurements for complexes it is clear that a quantitative relationship between the reported activities (IC_{50}) and the observed flap movements is not present.

3.2 Introduction

AIDS still is a major health threat caused by a previously unknown agent that was reported in the mid-to-late 1970s[1-20]. Since the discovery that HIV is the causative agent of AIDS, extensive research on the development of new and improved therapies, targeting HIV, were conducted[1, 9, 21-23]. Human Immunodeficiency Virus type 1 protease (HIV-1 PR) was found to be one of the therapeutic targets[1, 9]. HIV-1 PR is classified as a C_2 symmetric homodimeric aspartyl protease with 99 amino acids from each monomer (see Figure 20)[9, 24]. The enzyme is essential for the virus's life cycle since it cleaves the *gag* and *pol* polyproteins into building blocks required for the formation of mature virions during its replication[9, 24]. The structure of the protease consists of three domains, which are referred to as the active site cavity, terminal domain (dimerization domain) and the flap domain[9]. The active site consists of the two tripeptide Asp25-Thr26-Gly27 units found at the interface of the core domain formed by the two monomers and is essential for dimer stabilization and the catalytic mechanism[9]. Ligand entry to access the active site is regulated by the two glycine-rich β-hairpin turn structures called flaps[9, 25-30]. The flap domain consists of amino acid residues from 43-58 in each subunit[25-27, 30, 31].

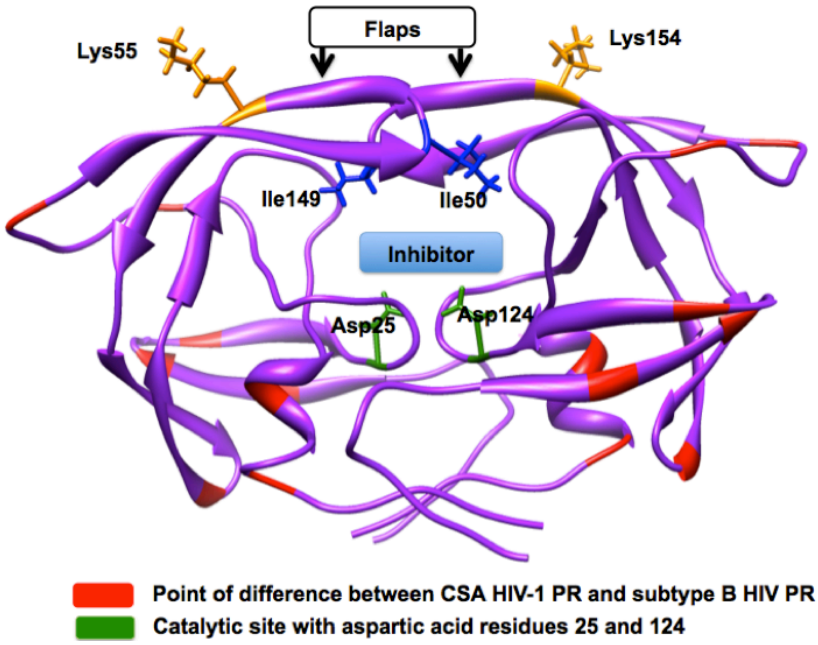


Figure 20. Structure of C-SA HIV-1 PR

Various computational and experimental studies such as MD simulation[12, 23, 25-30, 32-40], NMR[30, 31, 41] and X-ray[30], have investigated the flap-opening and closing mechanism of the HIV-1 protease subtype B for the ligand-bound PR and unbound enzyme[12, 25-30, 42]. It was reported that upon binding of the ligand to the HIV-1 PR, the flaps are pulled inward to tighten around the active site[26, 27, 30]. In the case of the free enzyme, however, the flaps are relatively distal from the active site but still substantially cover the catalytic triad and remain in contact with each other[26, 27, 30]. In most reported studies, two distinctive flap conformations are commonly observed; these are the closed form for the ligand-bound PR and the semi-opened form for the free enzyme[26, 27, 30, 43]. Several computational studies have intended to provide more insight into HIV-1 PR flap opening dynamics using different simulation protocols. Hornak *et al*. found, in unrestrained implicit solvation MD simulations in Amber with ff99 force field for the protein, that the open form of the free enzyme was in dynamic equilibrium with the closed and semi-open forms[26, 27, 30]. Toth and Borics employed MD simulations using Imagiro molecular mechanics program and OPLS-AA force field with GB/SA implicit solvent model to explore the mechanism of flap motion for the free and bound enzyme[28, 29]. They found that, during the simulation, the substrate induced the closing of the flaps into the closed conformation in an asymmetrical way through a hydrophobic intermediate state cluster. The same authors proposed that a network of weakly polar interactions between the flaps seems to be responsible for the stabilization of certain flap conformations over the others. Scott and Schiffer reported another unique feature of flap motion using MD simulations found in GROMOS96 package and the 43A1 force field with the SPC/E water model[25]. The results showed that the flaps curl in and burry the hydrophobic tips, Ile50 and Ile149, against the hydrophobic inside wall of the active site groove, to allow enough space for the inhibitor to access the active site[25].

Different metrics have been proposed to describe the various flap conformational states of HIV-1 PR. The most commonly used one is based on the distance between the α-carbons of Ile50 and Ile149 residues[26-30, 44, 45]. This distance was observed to be 5.8 Å, 4.3 Å and 30 Å for the closed, semi-open and fully-opened conformations, respectively[26, 27, 30]. Another parameter based on the distance between the α-carbons of K55 and K154 residues was introduced by Karthik *et al.*[12] following the

report by Galiano *et al.*[46] and Torbeev *et al.*[47]. The MD simulations were performed using Amber 9 with a ff99SB force field. According to this metric, the distances of 28 Å, 34 Å and 43 Å represent closed, semi-opened and fully-opened forms, respectively[12]. Recently, estimation of the Ile50 - Asp25 or the Ile149 – Asp124 distance was believed to be more realistic since it monitors the flap tip-tip (Ile50 - Ile149) distance, because the flap tip-tip distance can be affected by both flap tip curling and flap asymmetry[48, 49].

We have previously reported the first account of the flap flexibility of the unbound C-SA HIV PR[50], expressed in the laboratory of our co-workers. MD simulations in explicit solvation with ff03 force field were employed and the distance between the α-carbons of the flap tips (Ile50 – Ile149) were measured. These results demonstrated that the flaps of C-SA PR exhibited more flexibility than subtype B PR as the former is missing the E35-R57 salt bridge; which likely contributes to flap stability[50]. The first African HIV-1 subtype C PR (3U71.pdb) was solved at a 2.7 Å resolution[50]. It was found that subtype B and C-SA PR do not show significant difference in the overall tertiary folding. Crucial hinge-located variations have been identified at residue 36, (one for each part of the dimer) which are apparently closely related to E35-R57 salt bridge formation as seen in the C-SA PR and may be implicated in drug resistance via effects on flap dynamics.

In the current study, we aim to further explore the conformational changes of the flaps upon binding with different classes of pentacycloundecane (PCU) based cage peptide and peptoid inhibitors that were synthesised and tested for activity in our laboratory[14-17, 51-54]. To that end, we performed MD simulations for four independent models: (1) the free enzyme, (2) PCU-peptide/peptoid PR complexes, (3) ritonavir PR complex and (4) natural peptide inhibitor PR complex. The structures of the inhibitors and natural substrate that were used in this study are shown in Figure 21.

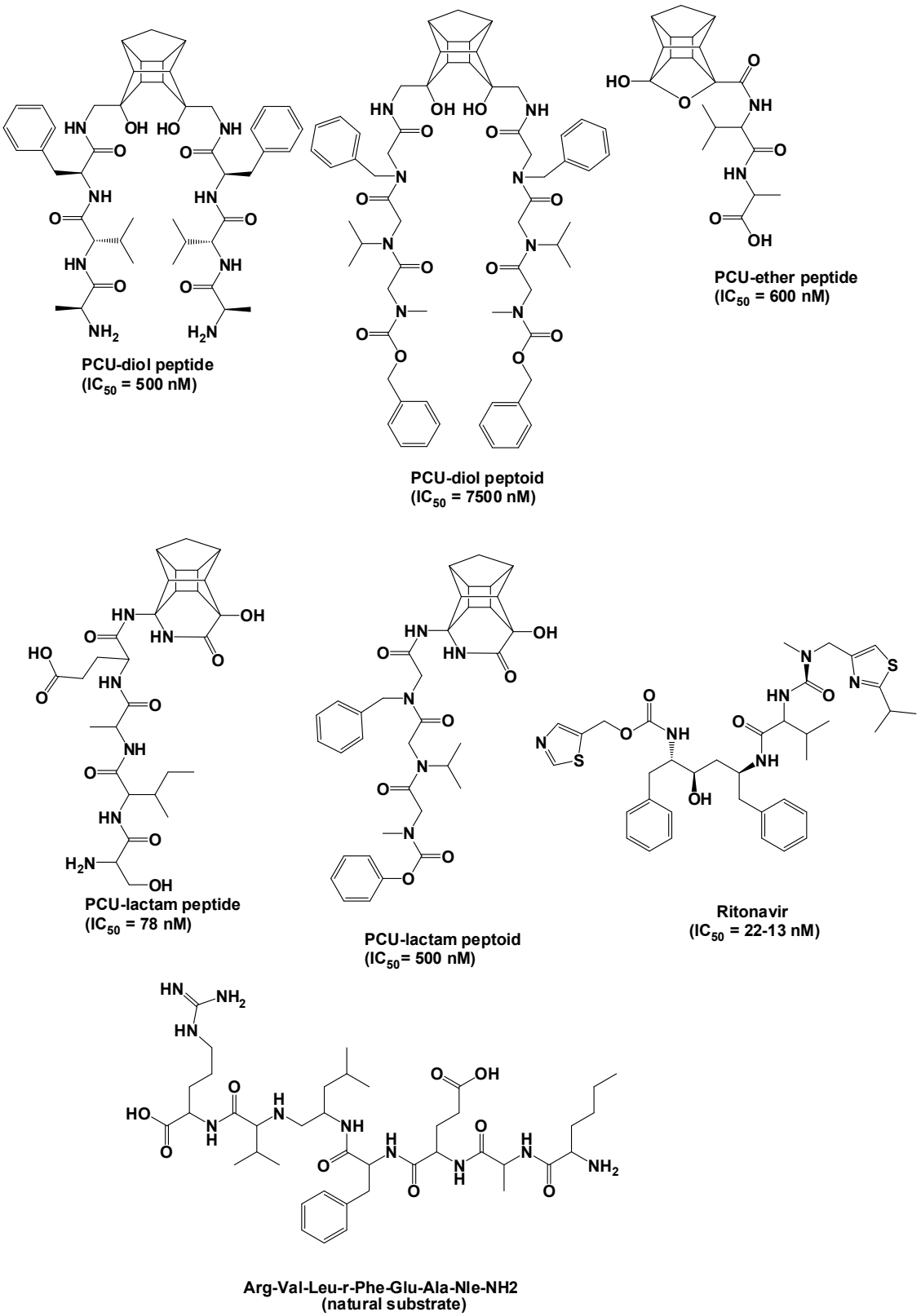


Figure 21. Structures of the five PCU-based compounds[14, 17, 51, 52, 54], ritonavir[55] and natural substrate[56] with their half maximal inhibitory concentration (IC_{50}).

We performed simulations (Amber software with the ff99SB force field) in both explicit and implicit solvents. Results were compared to previously published reports on subtype B HIV PR, where applicable. Moreover, we used several different metrics to determine the best descriptors to obtain optimal information about flap motion.
Comparison of the MD results for the four models will enable us to answer the question to what extent analysis of flap dynamics can be used as a measure of effectiveness of different HIV PR inhibitors. Furthermore, this study can potentially guide us towards the design of more active HIV PR inhibitors.

3.3 Material and methods

3.3.1 Computational Simulation

The following sections present an outline of the ligands and receptors used for molecular modelling. It also provides details about the software and the parameters that were employed.

3.3.1.1 Preparation of PCU-derived peptide/peptoid inhibitors and Ritonavir

Figure 21 presents the structures of the five PCU-based compounds[14, 17, 51, 52, 54], ritonavir[55] and natural substrate[56] model that were investigated in this study; the PCU–based compounds are peptides and peptoids. The PCU based HIV-1 PR inhibitors were constructed and optimized with the molecular mechanics optimization force field implemented in Hyperchem 6.0 software[57]. The ionization states of the aliphatic amine and carboxylic acid groups of the PCU-based inhibitors were reported to have no significant effect on the docking results when they are either in the neutral or ionized states[16].

Docking of the five neutral compounds against C-SA protease was performed. From the Protein Data Bank, 1HXW and 1A8K accession codes (subtype B HIV-1 PR) were downloaded, after which the enzyme was separated from the ligand (ritonavir - 1HXW)

and natural substrate (1A8K) by using USCF Chimera software[58]. The minimized PCU inhibitors were then subjected to docking studies to obtain a favourable inhibitor conformation within the PR active site.

3.3.1.2 C-SA HIV-1 PR enzyme model system

Since the X-ray structure of the South African HIV-1 protease subtype C[50] (C-SA PR) was not available at the time we started this work, a model for C-SA PR was constructed from the reported X-ray data of subtype B HIV-PR (PDB accession code 1HXW) [55]. C-SA HIV PR differs from subtype B at eight positions; T12S, I15V, L19I, M36I, R41K, H69K, L89M and I93L (see Figure 20)[59]. Details for the construction and the validity of the modelled C-SA HIV have been demonstrated before[16, 17, 51, 54, 59-61]. The X-ray structure of C-SA PR has in the meantime been reported[50] and subsequent comparison between our model turned out to be favourable. The crystal structure of C-SA HIV PR was aligned with the modelled structure of C-SA PR by using VMD software[62] (see Figure 22) and the calculated RMSD was 1.03 Å.

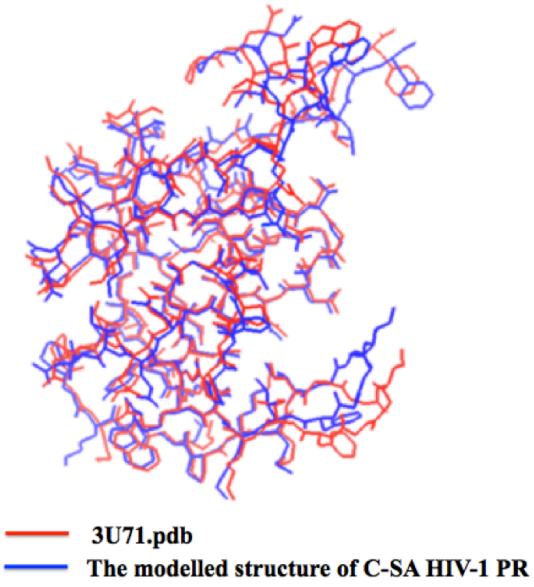


Figure 22. The structure showing the aligment of the 3U71.pbd and the modelled C-SA HIV-1 PR. This was achieved by using VMD software to obtain the RMSD.

3.3.1.3 Docking of the five PCU-based inhibitors, natural substrate and ritonavir into the active site of the C-SA HIV-1 PR model and subtype B

Before docking, all the PCU derived inhibitors were aligned and superimposed to ensure that all starting structures had the same 3D orientation with respect to the active site since this was found to improve the accuracy of the procedure. The docking approach adopted in this study has been fully described before[16, 54]. Ritonavir was not docked as the X-ray structure of the type B complex (1HXW) was used for subsequent MD studies. The X-ray structure of an analogue of the CA-p2 site (1A8K)[56] was used for MD analysis (see Figure 21) for the substrate structure (Arg-Val-Leu-r-Phe-Glu-Ala-Nle-NH_2)). Note that this is the same sequence (Ala-Arg-Val-Nle-*p*-nitro-Phe-Glu-Ala-Nle-NH_2) that was synthesised in our lab[15] and used for *in vitro* testing of the IC_{50} values of these inhibitors. The only small change introduced is the nitro group at the *para* position on the phenylalanine, in order to monitor the catalytic activity of the protease.

Docking studies were performed using the AutoDock software[63, 64]. During the procedure, Geisteger partial charges were assigned and AutoDock atom types were defined using the AutoDock Tool graphical user interface supplied by MGL Tools[65]. The docked conformations were generated using the Lamarckian Genetic Algorithm, which is the best method compared to simulated annealing method docking available in AutoDock[63]. The initial population was set at 150, the number of Genetic Algorithm cycles set at 10 and the maximum number of generated structures (conformations) at 27 000. The dimension of the grid parameter files defined in AutoDock Tools was X= 52 Å, Y= 40 Å and Z= 40 Å and that of the center grid box was and X=16.572 Å, Y= 24.503 Å and Z= 2.179 Å. All other parameters were left with the default values. The docked conformations of each inhibitor were ranked into clusters based on the binding energy. The program was set to automatically choose only the top ranked conformations, which were visually analysed and selected for MD simulation. Selection was based on the binding energy and whether a chemically sensible position of the inhibitor was present in the active site. The chosen docked complex structures are available as supplementary material.

3.3.1.4 Preparations of the PCU-based compounds, natural substrate, ritonavir and receptor for MD simulation

All complexes selected for the MD simulation were extracted from the docking log file (dlg) generated from the docking results for each investigated PCU-based peptide/peptoid inhibitors. Hydrogen atoms were added to the ligands, natural substrate and ritonavir with the Avogadro software[66] while keeping the amine and carboxylic groups in a neutral state. All hydrogen atoms were deleted from the receptor (subtype B and subtype C PR) since LEaP module in Amber 10[39] adds them. Visual inspection of each complex was performed to ensure that the correct structures were generated. The prepared systems were then used for MD simulation in Amber 10[39]. In the current study deprotonated states for the Asp25/Asp124 of the PR were considered since it was reported that docking of the C-SA complexes where either Asp25/Asp124 were protonated, gave similar docking behaviour to the unprotonated Asp25/Asp124 residues[15]. Also, MD simulations of the monoprotonated and deprotonated Asp25/Asp124 of the PR gave similar binding free energies[67].

3.3.1.5 Molecular dynamics simulations

A) Explicit solvation

The MD simulation was performed using the AMBER 10 software[39] with the AMBER force field, ff99SB. All-atom force field parameters of inhibitors were obtained using the ANTECHAMBER module and the Generalized Amber Force Field (GAFF)[68] and AM1 BCC charges[68] were used to obtain force fields parameters for the ligands. The LEaP module in AMBER added hydrogen atoms of the receptor. The protein was placed at the center of cubic box with TIP3PBOX 8.0 Å and approximately 8482 water molecules were added. Six chloride ions were randomly added to neutralize the system to balance the +6 charge of the HIV-PR. The particle-mesh Ewald method[69] was used to handle long-range electrostatic interactions. The system was partially minimized, first to avoid insufficient pre-equilibration by 2500 cycles of steepest descent minimization; this was followed by 1000 cycles of conjugated gradient minimization under harmonic restraints with force constant 10 kcal/(mol·Å2) to all solute atoms. A non-bonded cut-off of 12 Å was applied. Next, the water molecules

were allowed to move while the protein and the ligand were held fixed. The system was then fully minimized with no restraint for 200 steps. Canonical ensemble (NVT)-MD was then carried out for 30 ps, during which the system was gradually heated to 300 K for 30 ps. The Shake algorithm[70] was applied to constrain the bonds that contained hydrogen atoms. The time step of the simulation was 2.0 fs. Finally, the 50 ns isothermal isobaric ensemble (NPT)-MD simulation was applied to the simulations without any restraints. The temperature was regulated at 300 K using the Langevin thermostat[71, 72] and the pressure was maintained at 1.0 atm using isotropic positional scaling. Trajectories were analysed every 1 ps using the PTRAJ module.

B) Implicit solvation

Aqueous solvation was also modelled implicitly by using the Generalized Born approach[73]. The MD simulation was performed with the AMBER force field, ff99SB. All-atom force field parameters of inhibitors were obtained using the ANTECHAMBER module and the Generalized Amber Force Field (GAFF)[68] and AM1 BCC charges[68] were used to obtain force fields parameters for the ligands. The LEaP module in AMBER added hydrogen atoms of the receptor. Six chloride ions were randomly added to neutralize the system to balance the +6 charge of the HIV-PR. The system was partially minimized, first to avoid insufficient pre-equilibration by 2500 cycles of steepest descent minimization; this was followed by 1000 cycles of conjugated gradient minimization under harmonic restraints with force constant 10 kcal/(mol·$Å^2$) to all solute atoms. A non-bonded cut-off was not applied. The system was then fully minimized with no restraint for 200 steps. Canonical ensemble (NVT)-MD was then carried out for 30 ps, during which the system was gradually heated to 300 K for 30 ps. The Shake algorithm[70] was applied to constrain the bonds that contained hydrogen atoms. The time step of the simulation was 2.0 fs. Finally, the 10 ns isothermal isobaric ensemble (NPT)-MD simulation was applied to the simulations without any restraints. The temperature was regulated at 300 K using the Langevin thermostat[71, 72] and the pressure was maintained at 1.0 atm using isotropic positional scaling. Trajectories were analysed every 1 ps using the PTRAJ module.

3.4 Metric parameters for estimation of flap movement analysis

In this study four different metric parameters (three distances and one dihedral angle) were used to measure the extent of flap opening; these are as follows: (1) the distance between the α-carbon atoms of I50 and I149 residues[26-30, 44], (2) the distance between the α carbon atoms of K55 and K154 residues[12], (3) the distance between the catalytic site and flap tip for each PR monomer[48, 49]. For monomer A, the distance was measured between the α-carbon atom of I50 and β-carbon (α-carbon) atom of D25 while the distance between the α-carbon atom of I149 and β-carbon (α-carbon) of D124 for monomer B was measured. (4) The angle formed from the α-carbons of residues G48-G49-I50 of monomer A and G147-G148-I149 of monomer B were measured to monitor the flap-curling behaviour. These metric parameters have been employed for the subtype B PR[12, 26-30, 44, 48, 49]. In this study these metric parameters were employed for both subtype B and C-SA PR and ligands complexed with C-SA PR for evaluating the extent of flap movement. It should be noted that metric (1) is known to generate misleading results with type B PR[12]. It was decided to nevertheless employ it for C-SA PR in order to determine if the same phenomenon is also observed.

The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) of the unbound PR and complexes were also measured. All snapshots were created with the Chimera software[58]. The distance, RMSF, RMSD measurements over the trajectories were generated using the Ptraj module of Amber 10[39]. Comparisons of the reported IC_{50} with MD studies for the complexes are discussed after defining all metrics.

3.5 Results and discussion

Achieving a molecular view of the enzyme movements in solution is difficult by any experimental methods. To provide such insight we performed MD simulations with explicit solvation starting from a closed form of the unbound C-SA HIV-1 PR. In order to measure the extent of flap movement, a metric parameter for measuring the distance between the α-carbon of Ile50 and Ile149 residues was employed. The distance

between Ile149 and Ile50 is shown in Figure 23 A; however, when employing this method, it was observed that there was no significant change in the enzyme movement in the MD simulations up to 52 ns. The change at about 9 ns is a small one (~3 Å), representing a slightly curled form of the flaps when compared with that observed for the implicit solvent simulations (for the implicit solvent model typical changes are up to ~28 Å. See Figure 23 B.

There were no major changes observed in the explicit solvation calculations and this method was computationally expensive and time consuming. Thus, it was decided to rather use an implicit solvation model for the remainder of this study.

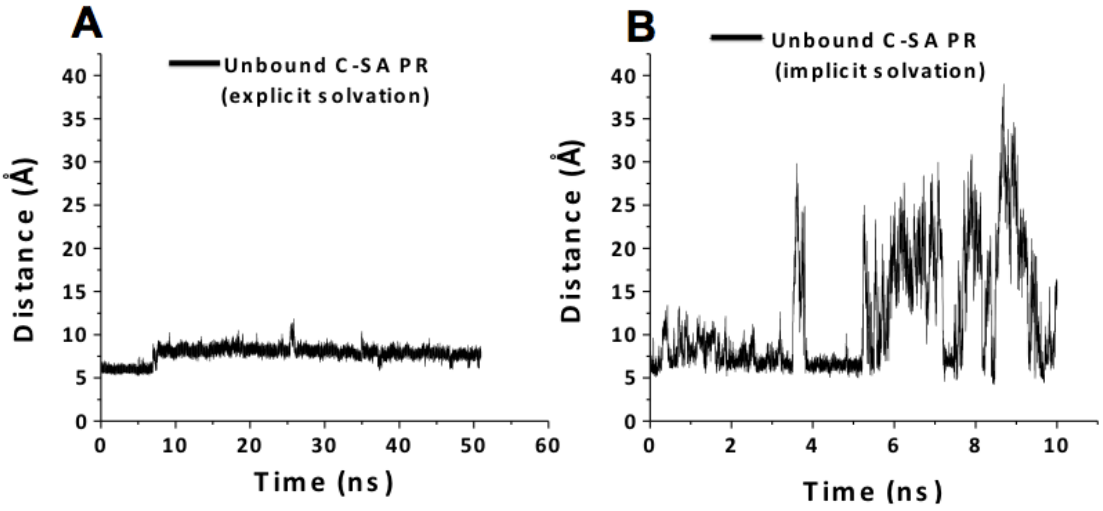


Figure 23. The evolution of the distance between the α carbon of residues I50 and I149 during the MD simulation in A (explicit solvation) and B (implicit solvation) of free C-SA PR.

A 10 ns MD simulation in implicit solvation was employed to investigate the C-SA HIV-1 PR flap motion. For comparison between subtype B and C-SA proteases the root-mean-square deviation (RMSD) for the α-carbon flap residues (43-58 residues) were plotted (see Figure 24); this measure provides the flap flexibility for these proteases.

In the HIV PR structure, residues 43-58 and 142-157 represent the flap region. The RMSD of flap A showed an average of 1.3682 Å and 1.3526 Å for subtype B PR and C-SA, respectively while flap B exhibited 1.2186 Å and 1.2906 Å for subtype B and C-SA PR, respectively. The RMSD plots represent deviations of the obtained conformation at a given time from the initial conformation; however, the similarity of the values makes it difficult to draw conclusions from the RMSD.

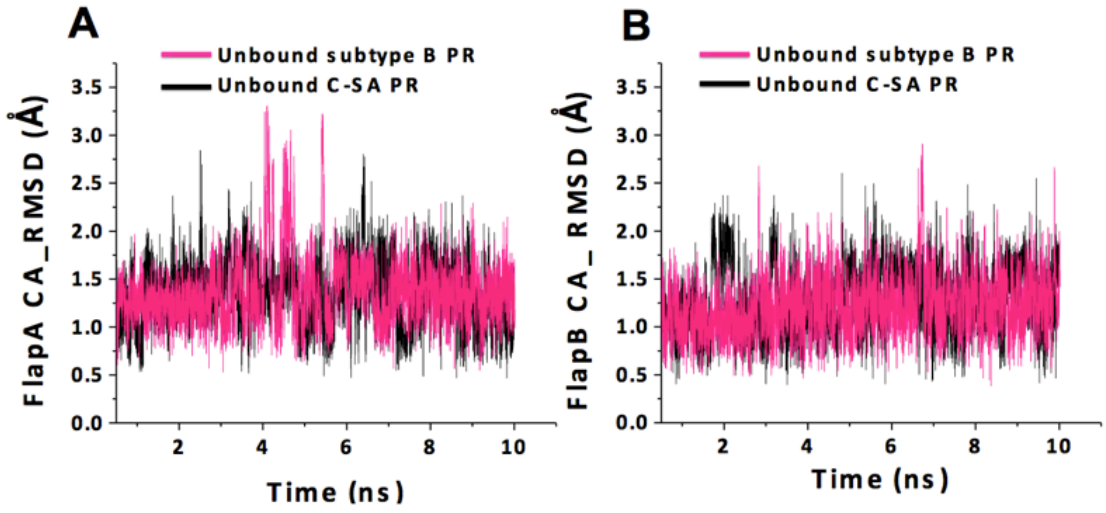


Figure 24. The RMSD of the Cα of the flap atoms of (A) residue 43-58 of monomer A and (B) residues 142-157 of monomer B (implicit water solvent).

The root mean square fluctuations (RMSF) of the proteases were plotted and shown in Figure 25 to give a better perspective on which residues are more flexible.

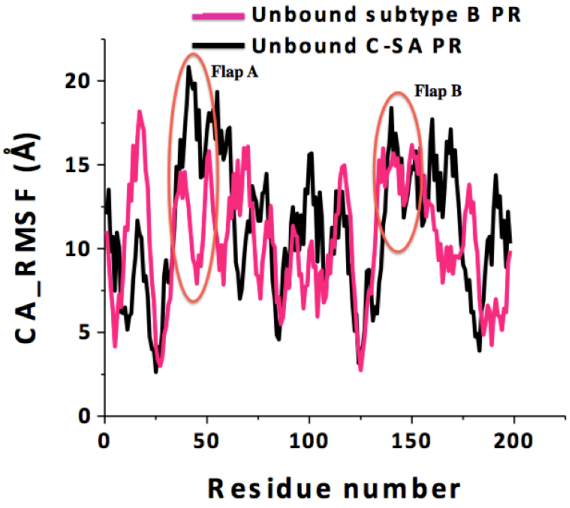


Figure 25. The RMSF for the subtype B and C-SA HIV-1 PR. Residues 43-58 and 142-157 represent the flap region for monomer A and B, respectively (implicit water solvent).

There is a significant difference between monomer A (Figure 25 A) flap residues for subtype B and C-SA PR. In contrast, monomer B (Figure 25 B) flap residues showed a smaller difference; this large change in the flap movements implies that C-SA PR exhibits greater flexibility than subtype B PR; a similar result was also reported by Naicker *et al*.[50].

In order to evaluate the effectiveness of the proposed metric parameters, visual inspection of the MD trajectory was performed to determine representative conformations of the respective flap movements during the course of the MD simulation. Snapshots of the subtype B PR are shown in Figure 26.

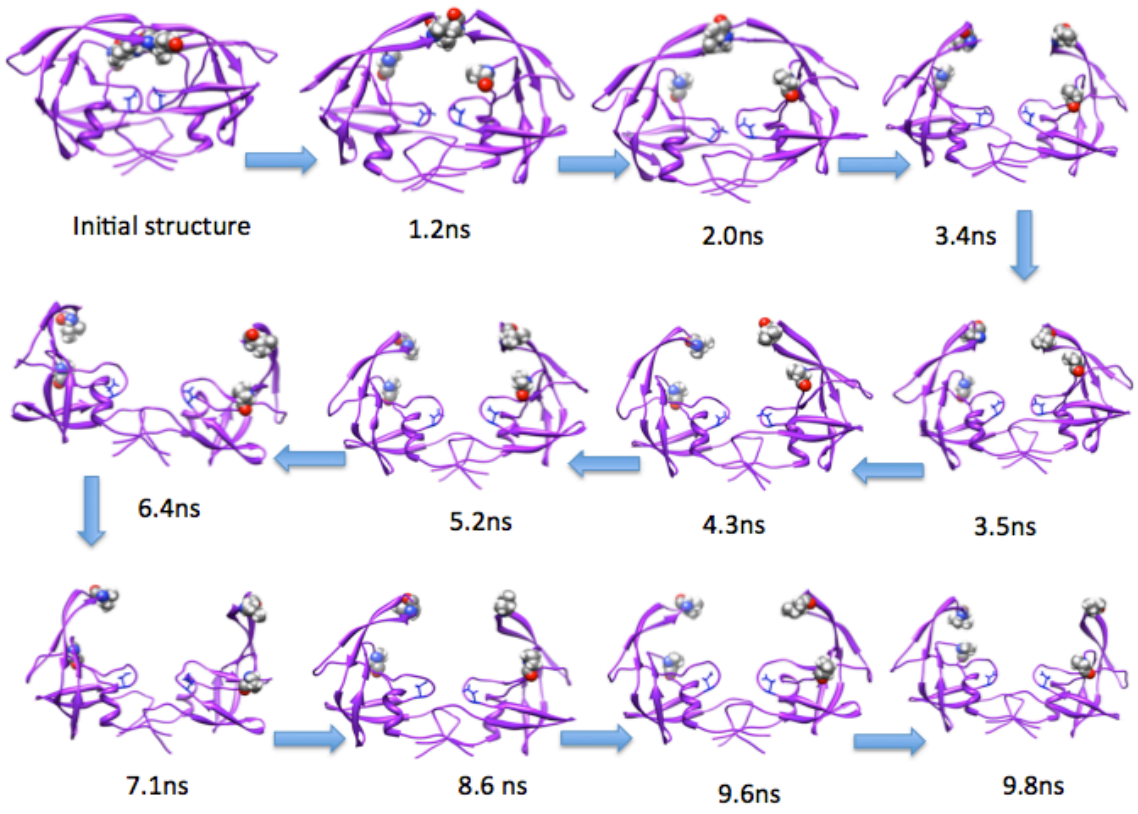


Figure 26. Snapshots of the unbound subtype B HIV-1 PR (implicit water solvent).

An initial closed form of the subtype B PR was observed which opened at 3.4 ns and remained in this form for the entire simulation. This shows that the subtype B flaps are essentially fixed in the opened form during the whole MD simulation.

Figure 27 shows snapshot views of the C-SA PR which demonstrates that the enzyme exhibits different structural conformations. The initial conformation was closed. At 1.2 ns a curling form was observed which was also previously reported, by Scott *et al*.[25]. Curling precedes flap opening. In our study, one flap remained in the semi-opened form while the other flap was curled towards the P1 loop (residues Pro79, Thr80 and Pro81). A semi-opened form of the protease was observed after the curling conformation (at 2.0 ns), which resulted in a opened conformation at 3.5 ns. The flaps returned to the semi-opened conformation at 4.3 ns. At 5.2 ns a fully-opened conformation was observed and the flap separation (the Ile149-Ile50 distance) reached a maximum of 28 Å at 8.6 ns.

These results indicate that C-SA PR is much more flexible than subtype B PR.

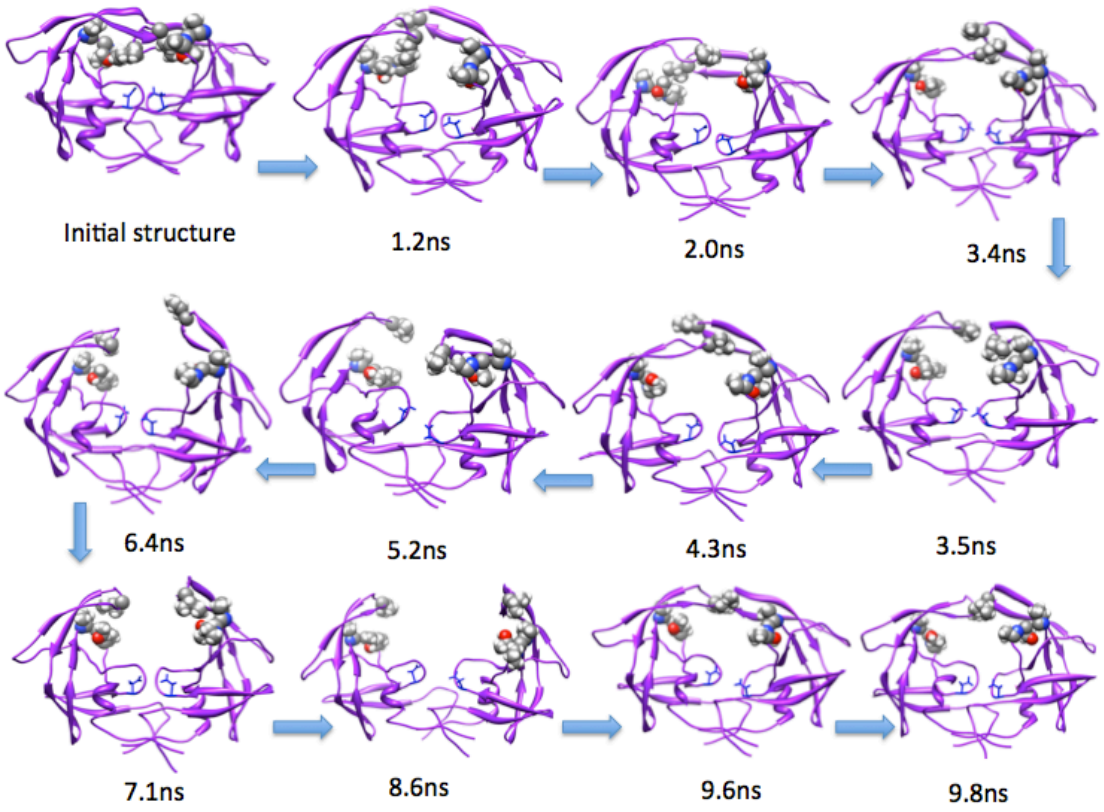


Figure 27. Snapshots of the unbound C-SA HIV-1 PR (implicit water solvent at 300K).

Evaluating the distance between the flap tips (between the α-carbon of I50 and I149) as suitable flap dynamics metric.

Unbound C-SA HIV PR

Karthink *et al*. reported that measurements of the distance involving the α-carbon atoms of residues Ile50 and Ile149 to monitor flap movements are misleading since it shows a contradictory trend in the distance of the semi-open form (4.3 Å), closed-form (5.8 Å) and opened form (>10 Å) of the unbound subtype B PR[12]. In this study it was found that the distances for the closed, semi-opened and opened forms of unbound C-SA were ~6 Å, ~10 Å and >15 Å, respectively; that is, the unbound C-SA PR (Figure 27) gave a

different trend from the one for subtype B reported by Hornak *et al.* [26]. An example of misleading results from this metric is a false positive result at the distances around 10 Å, which suggest that the flaps were open. On closer inspection the enzyme structure represents flap-curling or a semi-opened form.

This metric parameter was also used to probe the flap movement of the inhibitor/enzyme complexes.

Inhibitor bound C-SA PR

In the current study ritonavir was used as a reference inhibitor since it is an approved drug for HIV/AIDS treatment targeting the protease enzyme. The flap movements of five PCU-based inhibitors, PCU-diol peptide, PCU-diol peptoid, PCU-ether peptide and PCU-lactam peptide and PCU-lactam peptoid (shown in Figure 21), complexed C-SA PR were compared to the ritonavir complex and the unbound enzyme (see Figure 28). The Ile50 and Ile149 distances for the ritonavir/C-SA PR complex and the unbound C-SA PR are shown in all plots for direct comparison. All the plots in this study are presented in this fashion. Ritonavir was found to adopt a closed conformation (see Figure 28) from 0 to 6 ns with distances of less than 10Å; however, after 6 ns the conformation constantly changed from the closed to the semi-opened form (about 10 Å). This semi-opened conformation was due to the interaction of the substrate with the flaps.

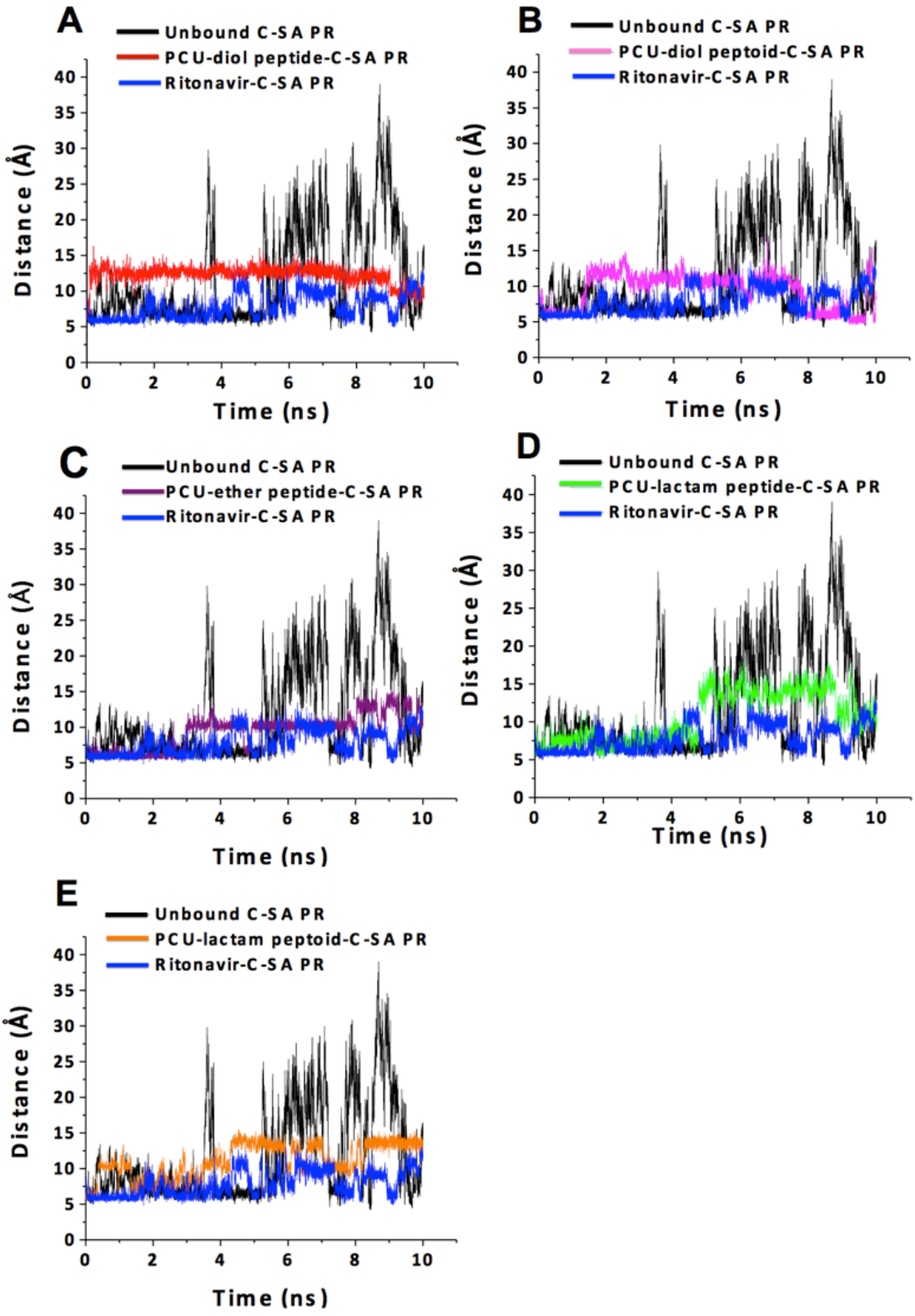


Figure 28. The evolution of the distance between the α-carbon atoms of residues Ile50 and Ile149 (implicit water solvent, 300K). The peptide/peptoid structures are presented in Figure 21.

For the PCU based compounds (Figure 28 A-E) it was observed that there is a significant difference between the Ile50 and Ile149 distances for the complexes and the unbound C-SA PR; for all the investigated cage compounds there is a reduction of the distances compared with the unbound C-SA HIV-1 PR. It is clear that the ritonavir/enzyme complex is tighter than the PCU-diol peptide complex (Figure 28 A) while PCU-diol peptoid (Figure 28 B) shows similar distances to the ritonavir/enzyme complex at 0-2 ns and at 6-10 ns. Analysis of the flap movement shows that PCU-ether peptide and PCU-lactam peptide compounds (Figure 28 C and D) behave similar to ritonavir, although it is clear that ritonavir forms a tighter complex. The other PCU-lactam peptoid (Figure 28 E) behaves similarly at the first 4 ns but slightly differently (show larger distances) for the rest of the simulation. These results suggest that in all cases that ritonavir (IC_{50} = 22–13 nM) forms tighter complexes than the PCU-based complexes. The PCU inhibitors that induce tighter complexes according to this metric are the PCU-ether peptide (IC_{50} = 600 nM), PCU-diol peptoid (IC_{50} = 7500 nM) and PCU-lactam peptide (IC_{50} = 78 nM) (Figure 28 B, C and D).

Natural substrate-bound subtype B and C-SA PR

From the measured distances in Figure 29 A, it was found that the ritonavir and the natural substrate complexed with subtype B gave relatively similar distances (about 10 Å) for the entire simulation. In some cases the natural substrate complex exhibited slightly lower distances (above 6 ns).

Figure 29 B, shows a comparison of the Ile50 and Ile149 distances for ritonavir and the natural substrate complexed to the C-SA PR. It is clear that both complexes behave quite similarly, although on average the natural substrate induces slightly more tight complexes.

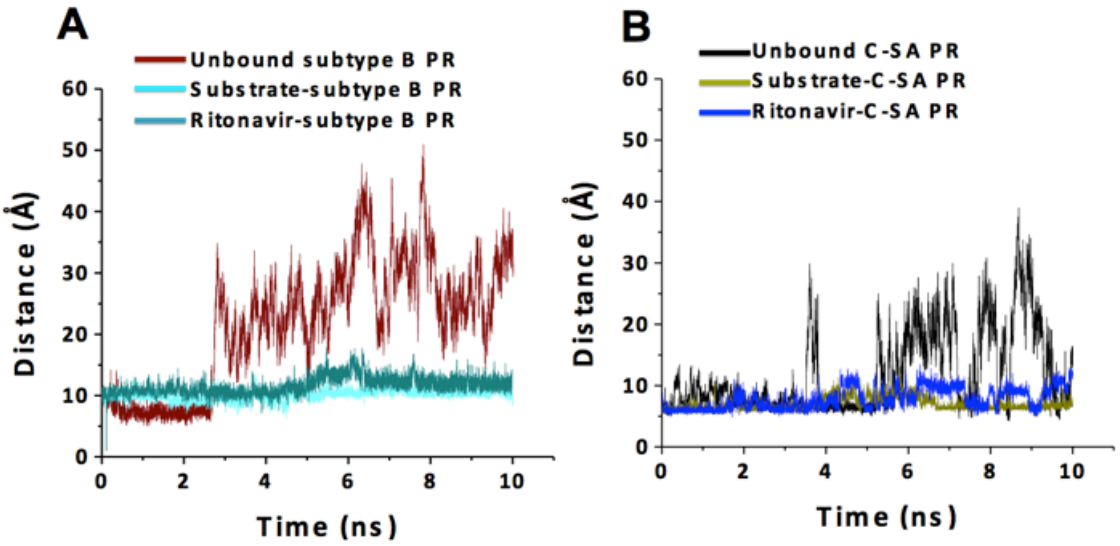


Figure 29. The evolution of the distance between the α-carbon atoms of residues I50 and I149 in the subtype B PR (A) and C-SA PR (B) (implicit water solvent, 300K).

Also the complexes for subtype C-SA show a tighter system than for subtype B. For C-SA, ritonavir and the natural substrate both gave distances around 6 Å during the first 4.5 ns, which represent a closed form. However, after 4.5 ns, the ritonavir complex interchanged its conformation between the closed and semi-opened form and this was due to interactions with the C-SA PR flaps more than the active site. The substrate is held and fixed in one conformation while the ritonavir complex shows interaction (close contacts) with the flaps.

Evaluating the distance between the flaps (Cα K55 and Cα K154) as a suitable flap dynamics metric.

Unbound C-SA PR

Karthik and his colleagues introduced a new parameter for monitoring the flap opening by measuring the distance of the α-carbon atoms of residue K55 and K154. When using this parameter for subtype B, it was observed that the distance of the closed form was 28 Å, the semi-opened form 34 Å and the opened form 43 Å[12]. When employing the same metric parameter introduced by Karthik *et al.*, it was found that the distance

between the α-carbon of K55 and K154 of the unbound C-SA PR was ~23 Å, ~ 30 Å and >35 Å for closed (at ~ 0-1 ns), semi-opened (at ~ 2-3 ns) and opened form (at ~ 3.5 ns and ~ 7 ns, and ~ 8-9 ns), respectively (see Figure 30). The values obtained in this study for C-SA PR were different from the ones proposed by Karthik *et al.* for subtype B, but both follow the same trend. This metric gives more information about the extent of flap opening however it does not show the flap tips behaviour that results into flap-curling.

Inhibitor bound C-SA PR

It was found that the PCU-based inhibitors and ritonavir complexes remained in the closed form, corresponding to approximately 23 Å or less. Figure 30 A-E shows the K55 - K154 distance evolutions for the PCU-based compounds. It was observed that the PCU-based compounds gave no significant difference in the distance evolution compared with the ritonavir complex. In comparison to the previously discussed parameter (distance of I50 - I149) it was found that these results do not conclusively describe the extent of the flap opening. Another parameter proposed by Seibold *et al.*[49] was the distance between the flap tips and the catalytic aspartic acid residue from each monomer which was one of the evaluated metric parameters in this study.

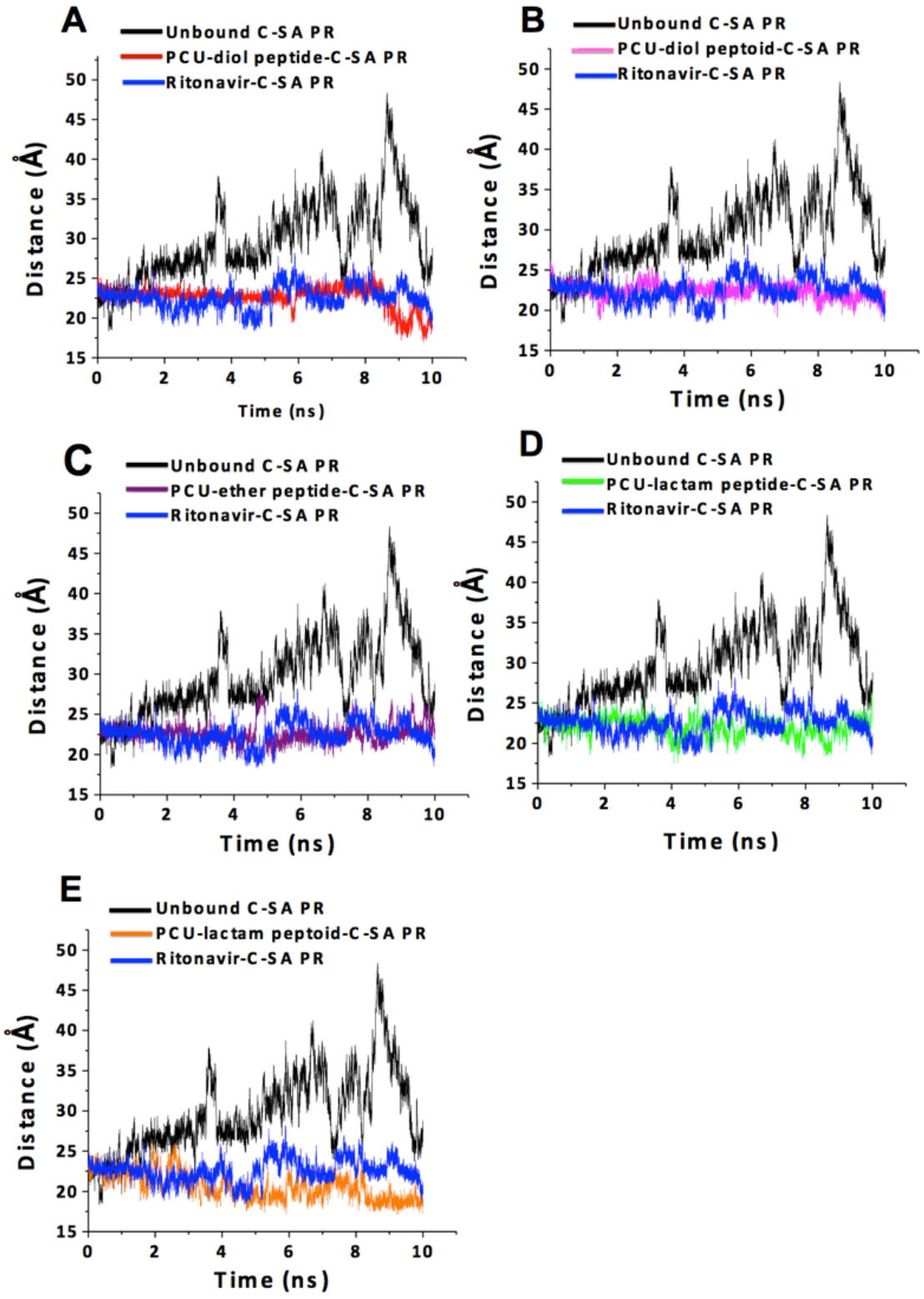


Figure 30. The evolution of the distance between the α-carbon atoms of residues K55 and K154 (implicit water solvent, 300K).

It is clear that the PCU based inhibitors behave very similarly to ritonavir. Since the latter is known to be considerable more active than the PCU derived inhibitors, it is apparent that analysis of flap movement during MD simulations is not sensitive enough to reflect qualitative comparisons between the strength of the interaction between the inhibitors and the enzyme.

Natural substrate-bound

For the subtype B complexes in Figure 31 A, it was found that both complexes showed similar distances for the first 5 ns. However, the natural substrate tends to exhibit larger distances of about 16 Å for the second 5 ns. In this case ritonavir forms a tighter complex during the last 5 ns in comparison to the natural substrate.

Figure 31 B shows the K55 - K154 distance for the C-SA PR. Distances of around 22 Å were observed for the entire simulation, which means that the substrate is tightly held in the closed form. This is in contrast to ritonavir, which showed fluctuations between 20-25 Å.

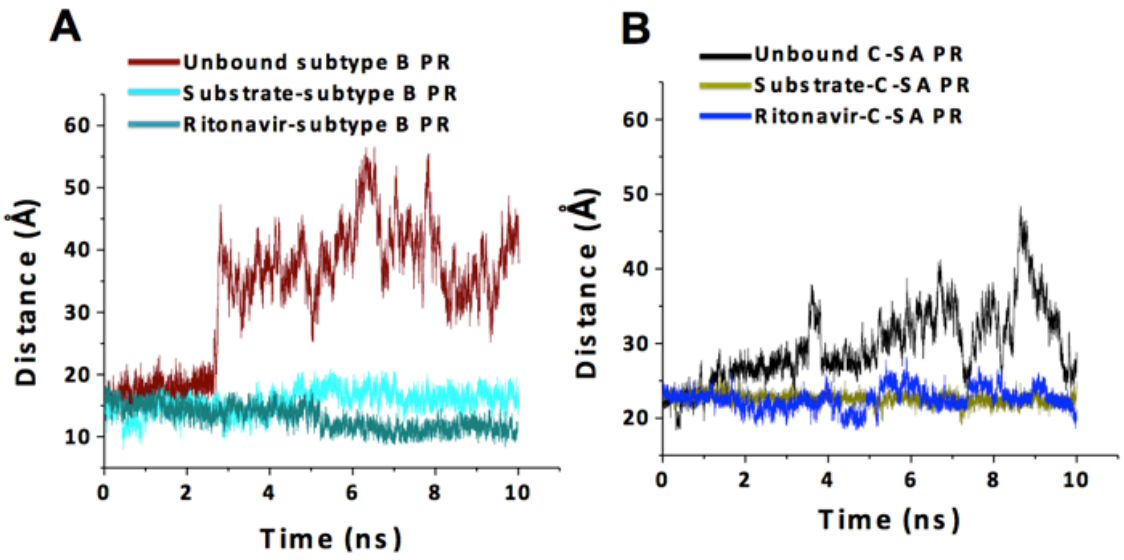


Figure 31. The evolution of the distance between the α-carbon atoms of residues K55 and K154 in subtype B PR (A) and C-SA PR (B) (implicit water solvent, 300K).

Evaluating the distance between the flap tips (I50 Cα or I149 Cα) and catalytic aspartic acid (D25 Cβ or D124 Cβ) for the PR as suitable flap dynamics metric.

Unbound C-SAPR

Seibold *et al*. also reported that measuring the distance between I50 and I149 is not the best parameter to monitor the extent of flap opening as it can be affected by flap-curling and flap asymmetry; therefore, it was proposed that the distance between the flap tip residues (I50 Cα or I149 Cα) and Asp catalytic residues (Asp25 Cβ and Asp124 Cβ) should be measured[74].

Figure 32 and Figure 33 show the distances for both the free enzymes and the inhibitor/enzyme complexes. It was found that the two flaps from each monomer behave differently. In the range of 0-4 ns, flap A of monomer A (I50 Cα - Asp25 Cβ) fluctuated between 19 Å and 31 Å (see Figure 32) while for flap B of monomer B (I149 Cα - Asp124 Cβ) the fluctuations increases from 10 Å to 28 Å with time (Figure 33) in the unbound C-SA PR. These results show that flap A is far from the catalytic site while flap B is close to the catalytic site. For both flaps the distance between the flap tips and catalytic aspartic acid was observed to decrease to 15 Å at ~4.5 ns and an increase to an average distance of 25 Å at 5 ns was observed. This shows that the flap movements are asymmetric.

Inhibitor bound C-SA PR

In comparing the distance between the flap tips and the catalytic aspartic acid, it was found that flap A (monomer A) of the unbound C-SA HIV PR behaves differently to when it is complexed. Interestingly, it was observed that the monomers, A and B, also, behave differently in the complexes.

Figure 32 A-E show that these complexes exhibited similar interaction behaviour with monomer A, for the first 6 ns, since the distances between the catalytic site and flap tips were lower (~16 Å) than the unbound C-SA HIV-1 PR(~25 Å). This indicates that one

flap is pulled more inward, towards the catalytic site by hydrogen bonding than the other flap which is further away from catalytic site. In contrast, the measured distances for monomer B in the ritonavir/enzyme complex was relatively similar to the unbound PR (see Figure 33). The PCU-based compound/enzyme complexes (see Figure 33 A-E) showed relatively similar fluctuation as the ritonavir complex. However, in some cases, the PCU-based compound/enzyme complexes exhibit smaller distances which shows that flap B is pulled more toward the catalytic site. In Figure 33 B, the PCU-diol peptoid complex shows smaller distances after 1.8 ns for the entire simulation when compared with the ritonavir complex. This demonstrates that flap B of the PCU-diol peptoid is pulled more inward towards the catalytic site.

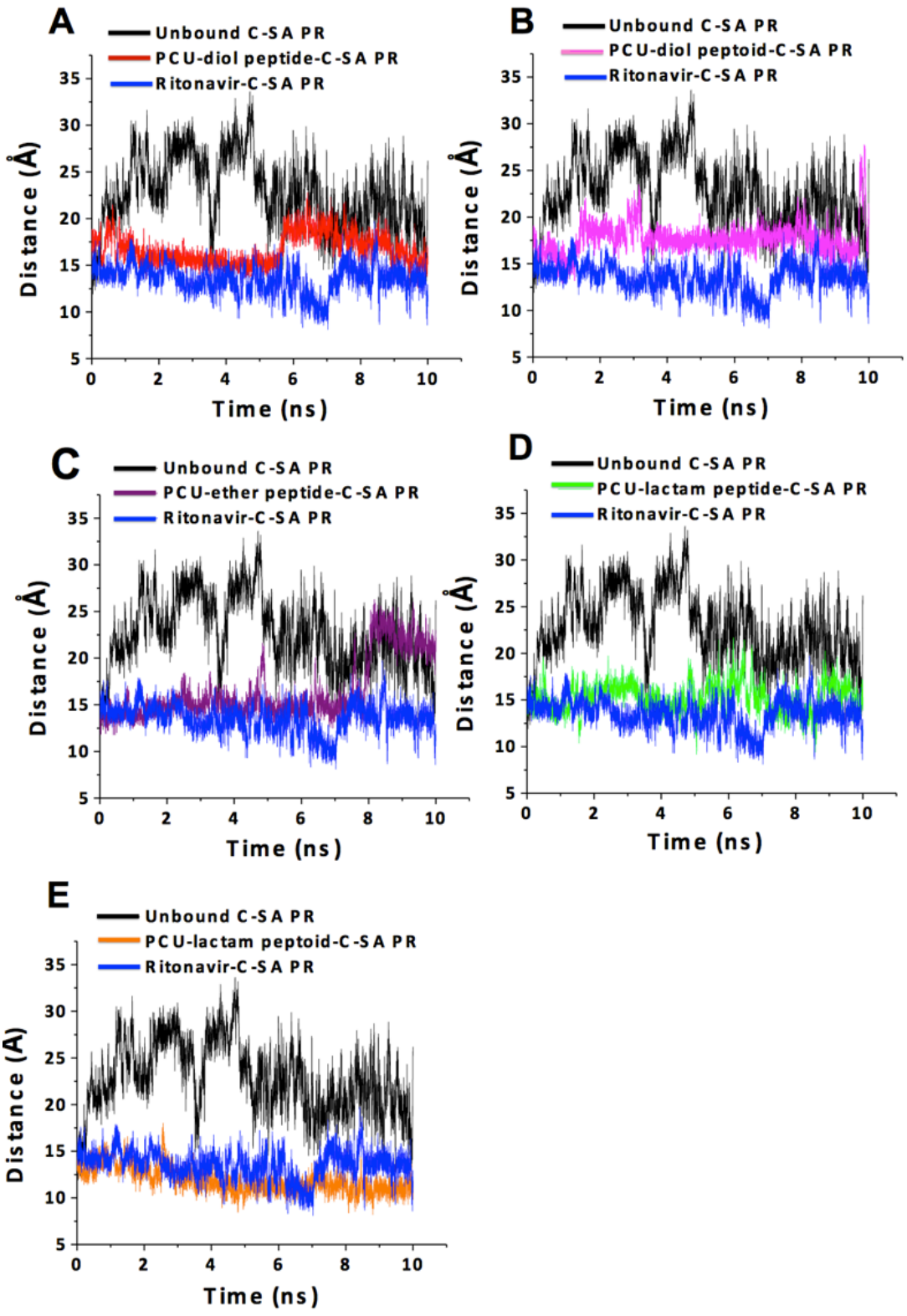


Figure 32.The evolution of the distance between the catalytic site and flap tip, for the β-carbon atom of residue D25 and the α-carbon atoms of residues Ile50 in monomer A in the PR (implicit water solvent, 300K).

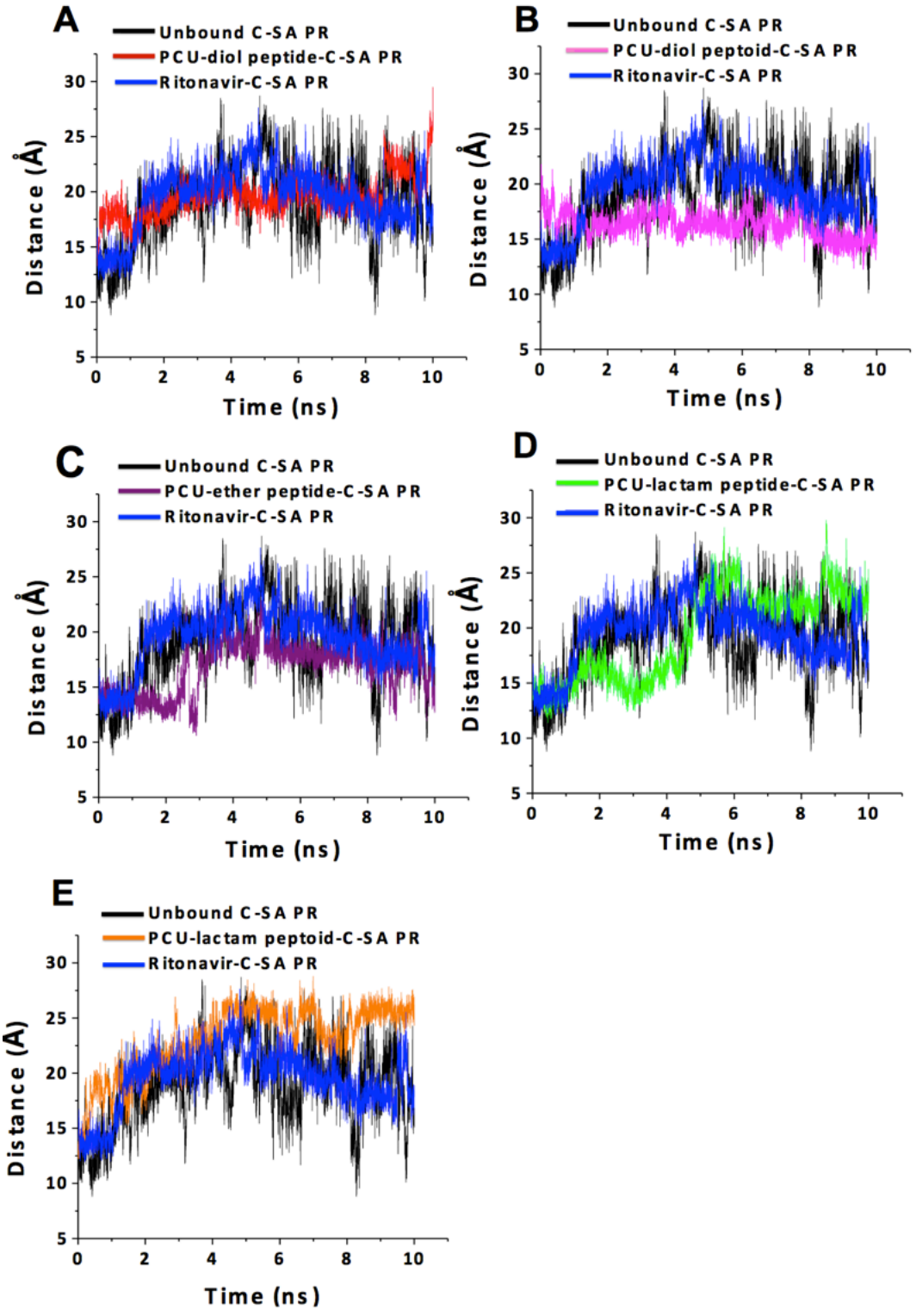


Figure 33. The evolution of the distance between the β-carbon atom of residue D124 and the α-carbon atoms of residues Ile149 in monomer B (implicit water solvent, 300K).

It is also apparent that with this metric parameter the analysis of flap movement during MD simulations is not sensitive enough to reflect qualitative comparisons between the strength of the interaction between the inhibitors and the enzyme.

Natural substrate bound to PR results

Figure 34 displays the distance between the catalytic aspartic residues and the flap tips, for subtype B PR (A) and C-SA PR (B). The natural substrate/subtype B complex exhibited shorter distances than ritonavir complex (Figure 34 A). The natural substrate is tightly held, thus, making the active site pocket smaller.

Conversely, ritonavir complexed with C-SA PR exhibited shorter distances than the natural substrate complex (Figure 34 B). This shows that the flap A of monomer A is pulled more inward than in the ritonavir complex, thus, making the active site pocket smaller.

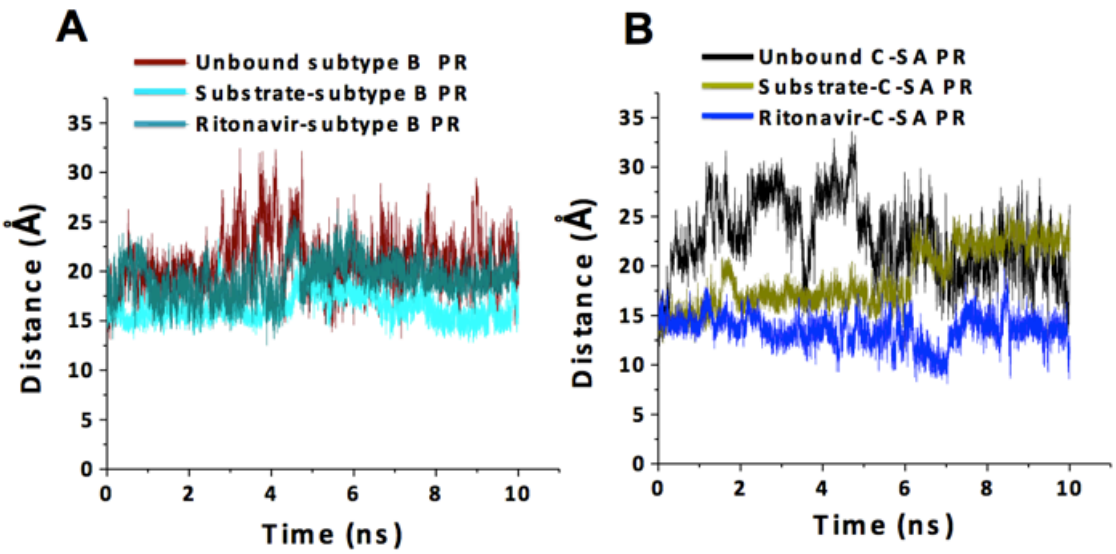


Figure 34. The evolution of the distance between the catalytic site and flap tip for between the β-carbon atom of residue D25 and the α-carbon atoms of residues Ile50 in monomer A in the subtype B PR (A) and C-SA PR (B) (implicit water solvent, 300K).

Figure 35 shows the evolution of the distance between the β-carbon atom of residue D124 and the α-carbon atoms of residues Ile149 in monomer B for subtype B PR (A) and C-SA PR (B).

The natural substrate and ritonavir complexes with subtype B PR exhibited similar distances (Figure 35 A) for the first 5 ns but between 5-8 ns the natural substrate showed smaller distances than the ritonavir complex. The last 2 ns for both complexes showed the same distances.

In Figure 35 B, flap B of monomer B in the natural substrate complex showed much smaller distances (about 12 Å) than for the ritonavir complex which tend to exhibit same distances (about 20 Å) as the unbound C-SA PR. In this case, when the natural substrate is complexed to C-SA HIV-1 PR the more tightly, the closed form is observed. It is possible that this is one of the reasons for the poor response when ritonavir is given to patients with the subtype C strain.

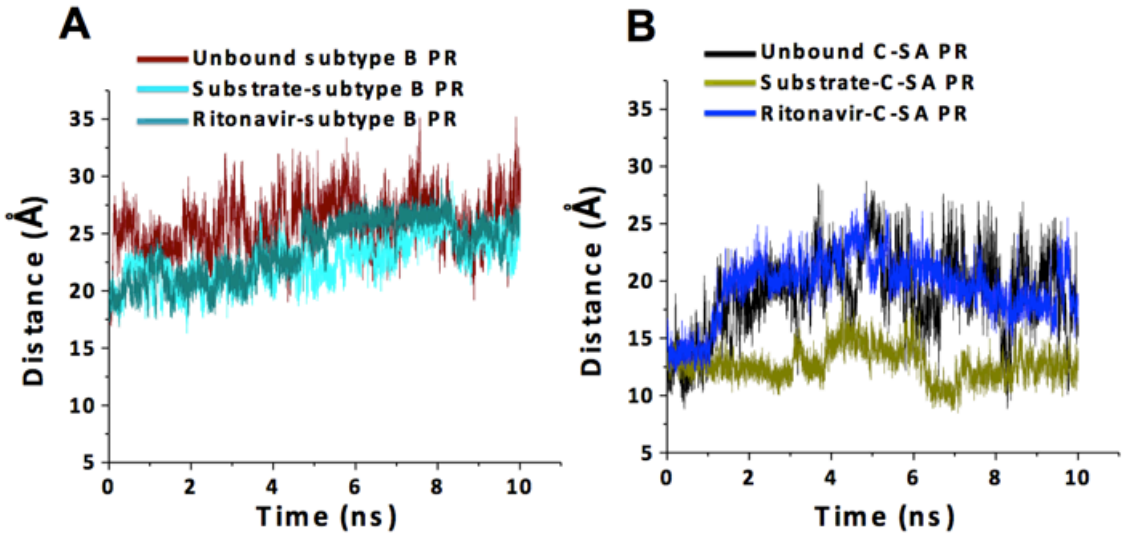


Figure 35. The evolution of the distance between the β-carbon atom of residue D124 and the α-carbon atoms of residues Ile149, in monomer B for subtype B PR (A) and C-SA PR (B). Simulations were performed in implicit water solvent at 300K.

RMSD for the backbone α-carbons

Figure 36 shows the RMSD of all backbone carbons. This, therefore, gives a measure of the flexibility of the protein as a whole. Similar to the plots generated by measuring the distance between the α-carbon residues of K55 and K154, the backbone RMSD plots showed the same trend for the complexes and for the unbound C-SA HIV PR. The unbound C-SA PR showed greater flexibility (greater than 2 Å) throughout the entire simulation while the complexes exhibited less flexibility with the average RMDS value of 2 Å (Figure 36 A-E).

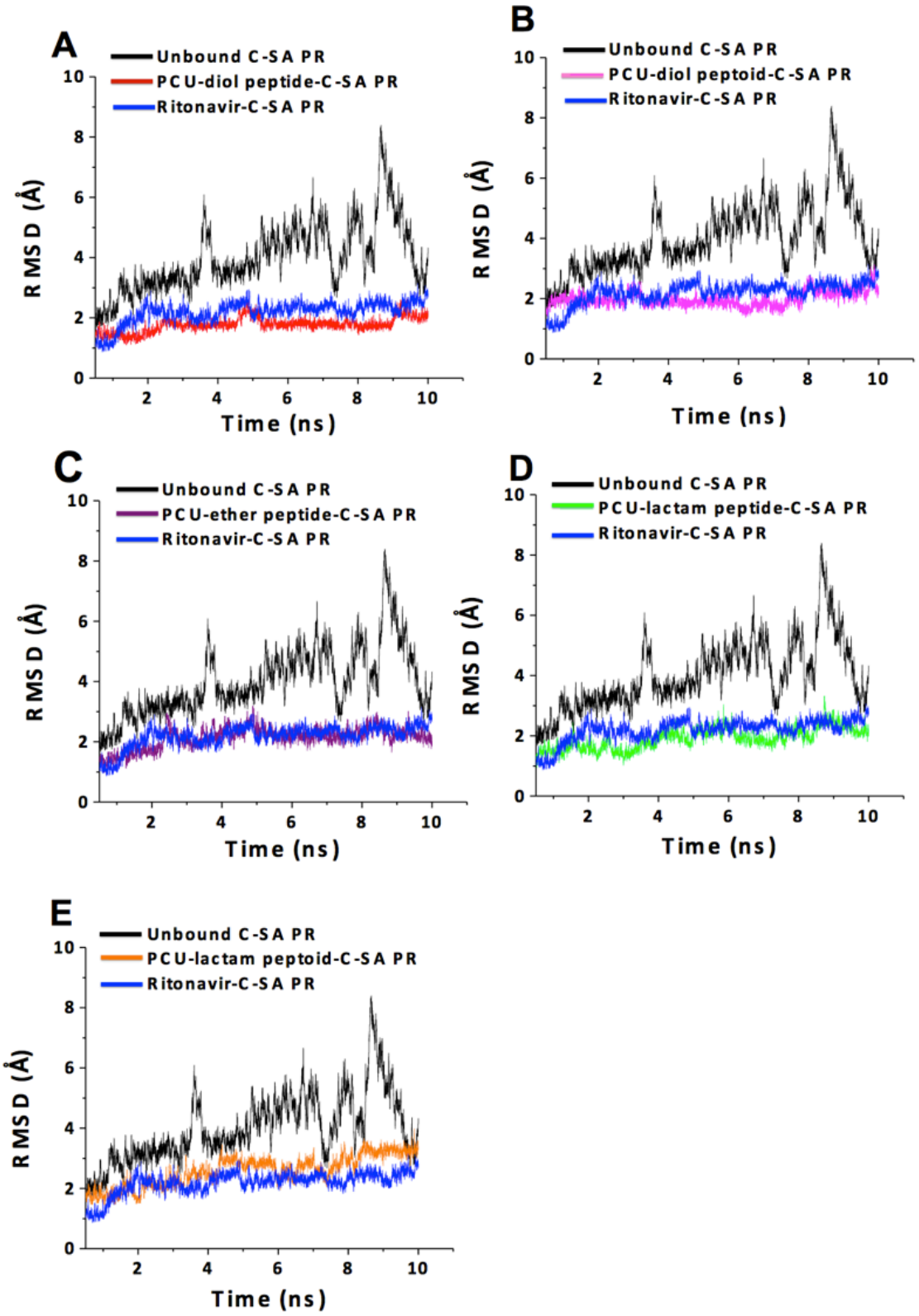


Figure 36. Backbone RMSD for the complexes and unbound PR (implicit water solvent, 300K).

Figure 37 A shows the RMSD for the natural substrate and ritonavir complexed with subtype B PR. Both complexes exhibit the same deviations in the first 4 ns; however, between 4 ns and 10 ns the natural substrate complex exhibited smaller deviations than the ritonavir complex.

Figure 37 B shows the backbone RMSD plots for the unbound, substrate bound and ritonavir bound subtype C PR. The natural substrate complex exhibited smaller deviations than the ritonavir complexed with C-SA HIV PR.

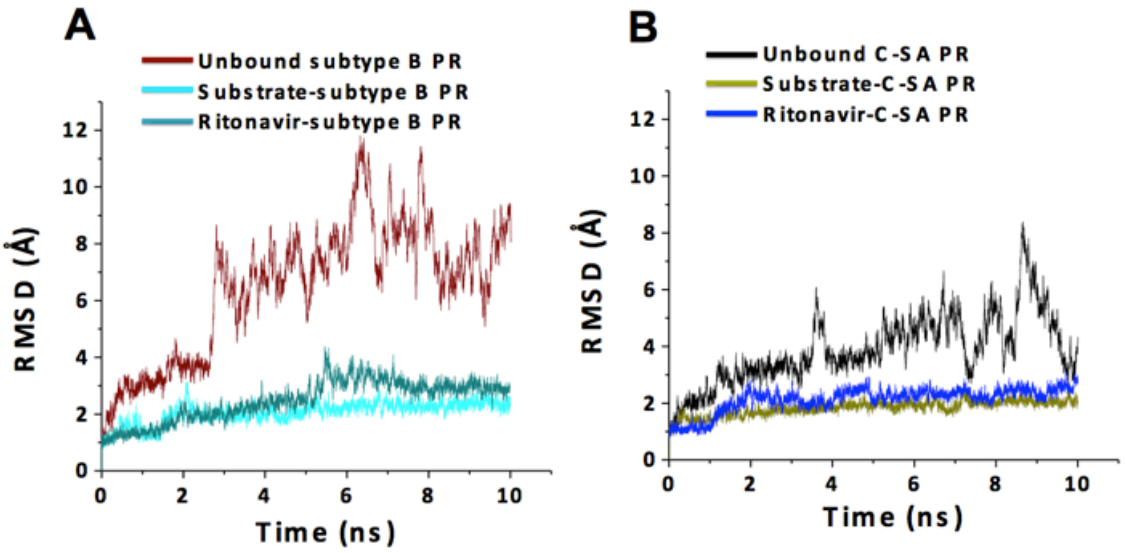


Figure 37. Backbone RMSD for the complexes and unbound subtype B PR (A) and C-SA PR (B) (implicit water solvent, 300K).

RMSF of the α-carbons of all residues

Unbound C-SA PR

The mobility of the flaps (residues 43-58 of flap A and 142-157 of flap B) in the unbound C-SA PR is greater than the ligand bound complex of all investigated ligands as shown by the RMSF of the α-carbon atoms (Figure 38 A-E). In the unbound PR dimer, flap A was more flexible than flap B, where the atomic fluctuation of the α-carbons reached 21 Å. The presence of a ligand reduces the flap flexibility of residues

43-58 (flap A) and residues 142-157 (flap B), thus, indicating a closed conformation of most of the complexes.

By comparison with all complexes, the PCU-diol peptide and PCU-diol peptoid (Figure 38 A and Figure 38 B) showed more flap movement while PCU-ether peptide and PCU-lactam peptide complexes were found to exhibit much lower flap fluctuation (Figure 38 C and Figure 38 D). This indicates that the flaps are held tightly by hydrogen bonding interactions to prevent flap fluctuations. On the other hand, the lactam peptoid (Figure 38 E) complex exhibited similar flap fluctuations as the unbound C-SA PR; this is probably a result of weaker binding affinity.

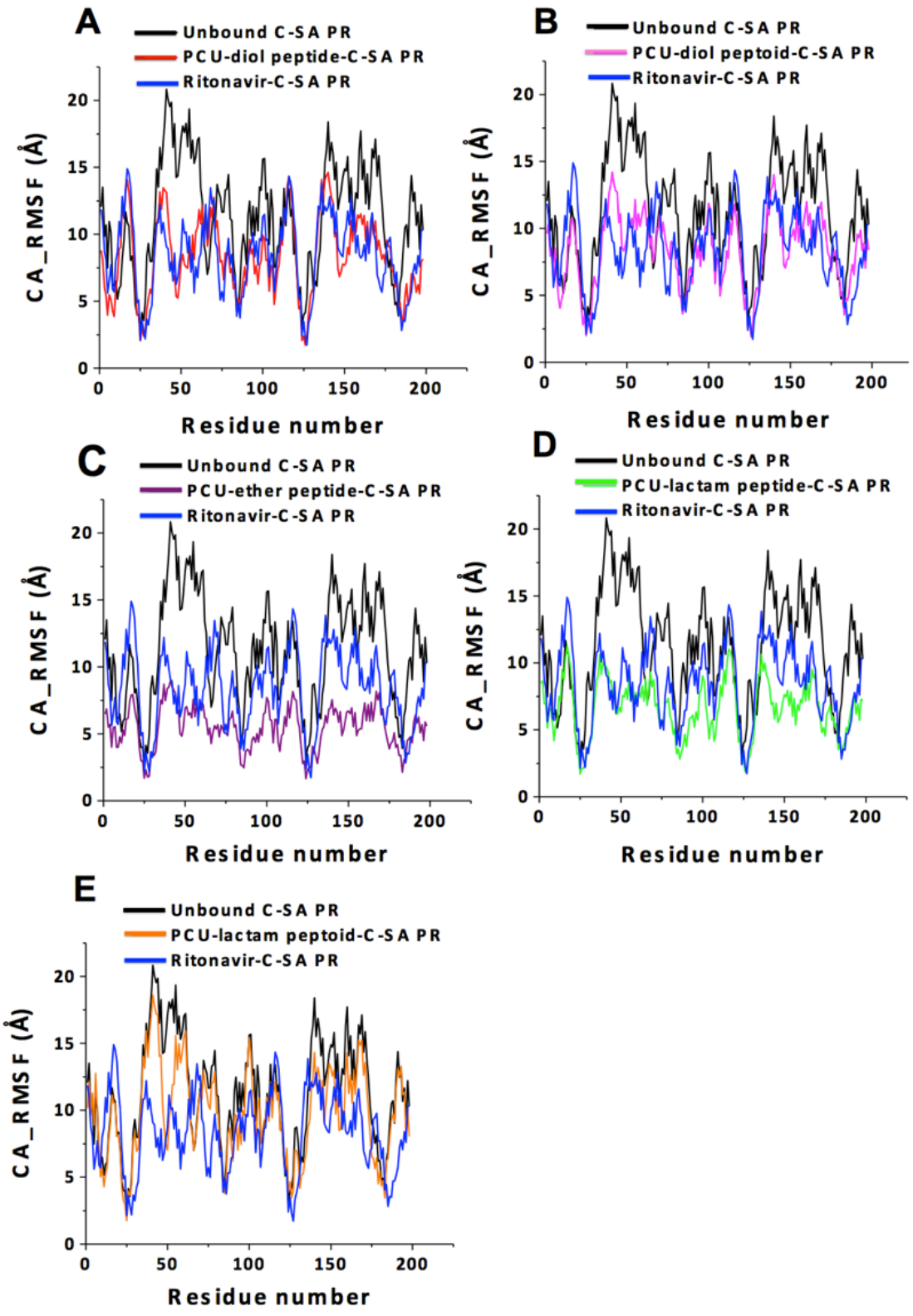


Figure 38. RMSF of the Cα atoms for the complexes and unbound C-SA HIV-1 PR (implicit water solvent, 300K).

Natural substrate-bound

Figure 39 shows the RMSF for the natural substrate bound to (A) subtype B PR and (B) C-SA PR. There is a reduction of RMSF for flap A and B for both HIV PRs. However the natural substrate bound to subtype B PR (Figure 39 A) seems to exhibit greater reduction in flap A (residues 43-58) than in flap B (residues 142-157). In Figure 39 B the ritonavir complex and the natural substrate complex showed relatively similar reduction of the flap motion.

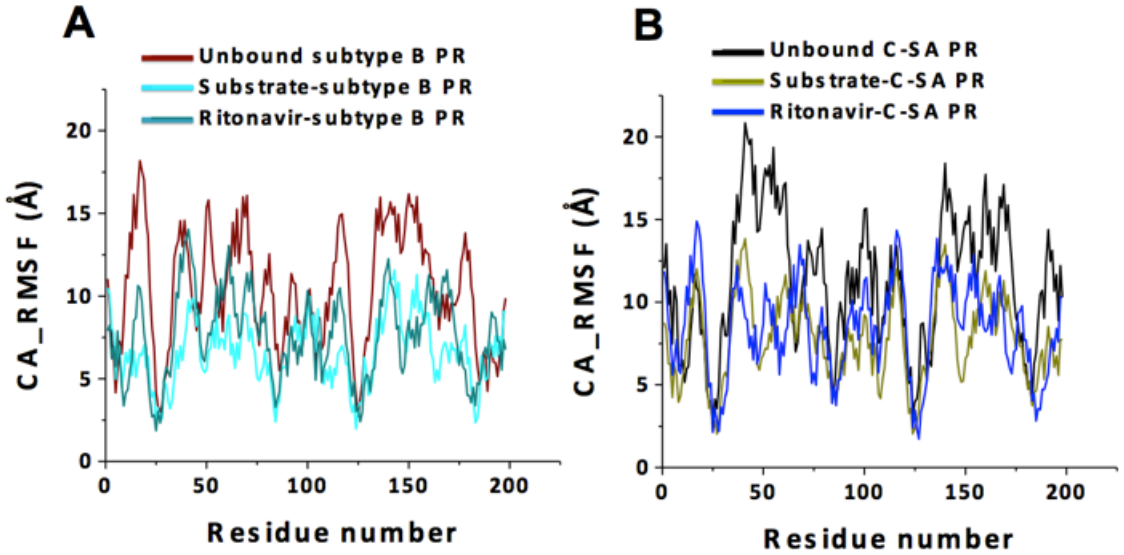


Figure 39. RMSF of the Cα atoms for the complexes, unbound subtype B HIV-1 PR (A) and unbound C-SA HIV-1 PR (B) (implicit water solvent, 300K).

Comparisons of unbound PR flap-curling with the curling of complexes

Scott *et al.*[25] observed, flap-curling, which was suggested to trigger flaps opening. During flap-curling, the angle formed from the α-carbons of residues G48-G49-I50 of monomer A and G147-G148-I149 of monomer B were proposed to monitor the flap-curling behaviour[49] (see Figure 40 A and B). The curling-in of the flaps results in the reduction of the angle from around 145° to 90°-120°. The lower angles represent the semi-open, curled or open form while angles above 120° represent the closed conformation. This parameter does not clearly define the conformation of the flaps or the extent of flap opening. The observed curling behaviour in the free enzyme seemed

to trigger flap opening however, in the complexes the curling event was due to the interaction between the flaps, the inhibitor and the P_1 loop (residues Pro79/Pro178, Thr80/Thr179 and Pro81/Pro180 from the protease). These interactions also contribute to the stability of the closed form and hence the entire complex.

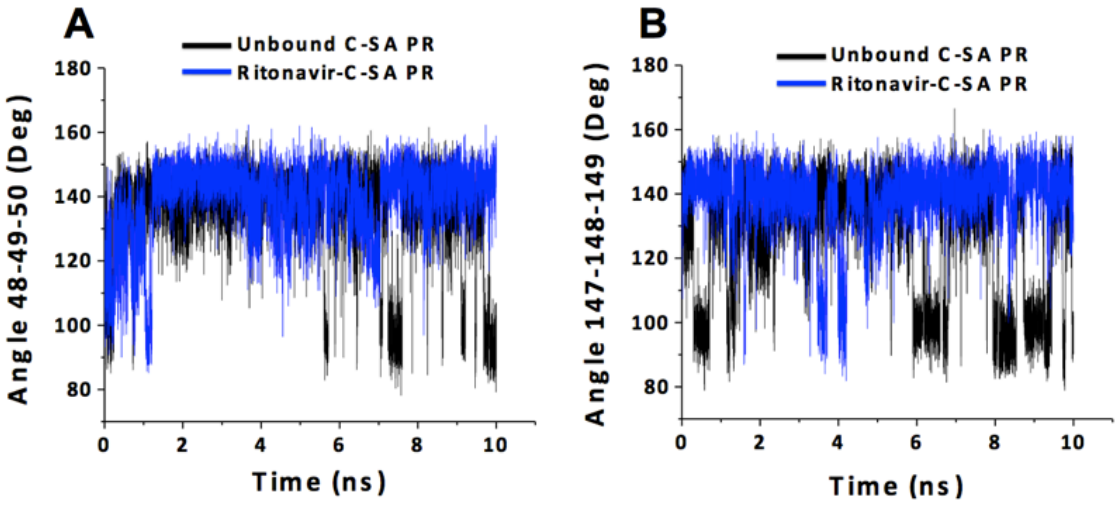


Figure 40. Flap-curling is defined by the angle formed from the three Cα of G48-G49-I50 in monomer A (A) and G147-G148-I149 in monomer B (B) versus time (implicit water solvent, 300K).

Comparisons between the inhibitors experimentally obtained IC_{50} with the observed MD results from the investigated matrices

The PCU-based compounds have been biologically tested ant their IC_{50} were shown in Figure 21. The PCU-lactam peptide showed the lowest IC_{50} of 78 nM while ritonavir exhibited[55] an IC_{50} of about 22-13 nM. When looking at the MD results and the flap movement measurements for the investigated PCU compounds it is clear that a quantitative relationship between the reported activities and the observed flap movements is not present.

Conclusion

To gain insight into the flap flexibility, MD simulations with implicit solvation were employed since shorter MD times explore the full conformational profile of the C-SA PR. Comparisons were made between the unbound and bound proteases. Comparisons between subtype B and C-SA PR showed that C-SAPR exhibits higher flexibility than subtype B; this is in agreement with the results by Naicker *et al.*[50]. Furthermore, it was found that the C-SA PR explored three forms during a 10 ns MD run (implicit water as solvent) *i.e.* a closed, semi-opened and opened form while the subtype B PR was dominated by the opened form. With reference to the measured parameters it was found that all complexes attained a closed form most of the time. This indicates that the presence of an inhibitor reduces the flap fluctuations, thus, preventing it from opening. The metric method for measuring the distance between the flap tips (Cα I50 – Cα I149) did not show entirely the extent of flap opening since it was affected by the curling of the flap tips which gave a false representation of flap opening while the flaps were in fact only curled. On the other hand, the metric for measuring the distance between the Cα K55 – Cα K154 showed complete flap movement while the distance between the flap tip and catalytic site (I50 Cα- Asp25 Cβ) and (I149 Cα- Asp124 Cβ) showed the tightness of the inhibitor in the active site. It was also found that measuring the dihedral angles [CAs of G48-G49-I50 (G147-G148-I149) monomer A (monomer B)] of the flap tips only shows which angles represent which conformations but not the extent of flap movement. Therefore, the best method to measure the extent of flap opening is by measuring the distance between Cα K55 – Cα K154.

3.6 References

1. A. Wlodawer, J. Vondrasek. Inhibitors Of HIV-1 Protease: A Major Success Of Structure-Assisted Drug Design. *Annual Review of Biophysics and Biomolecular Structure* 1998,**27**:249-286.
2. K. Bisetty, F. J. Corcho, J. Canto, H. G. Kruger, J. J. Perez. Simulated Anneling Study Of The Pentacyclo-Undecane Cage Amino Acid Tripeptides Of The Type [AC-X-Y-Z-NHMe]. *Journal of Molecular Structure: Theochem* 2006,**759**:145-158.
3. K. Covens, N. Dekeersmaeker, Y. Schrooten, J. Weber, D. Schols, M. E. Quinones-Mateu, A. Vandamme, K. Van Laethem. Novel Recombinant Virus Assay For Measuring Susceptibility Of Human Immunodeficiency Virus Type 1 Group M Subtypes To Clinically Approved Drugs. *Journal of Clinical Microbiology* 2009,**47**:2232-2242.
4. In: http://www.avert.org/aids-histroy-86.htm 04 April 2011.
5. In: http://www.rxlist.com/fuzeon-drug.htm 30 May 2011.
6. In: http://knol.google.com/k/hiv 20 May 2011.
7. A. Rodriquez, P. Mokoena, F. Corcho, K. Bisetty, J. J. Perez. Computational Study Of The Free Energy Landscape Of The Miniprotein Cln025 In Explicit And Implicit Solvent. *Journal of Physical Chemistry* 2011,**115**:1440-1450.
8. P. R. Batista, G. Pandey, P. G. Pascutti, P. M. Bisch, D. Perahia, C. H. Robert. Free Energy Profiles Along Consensus Normal Modes Provide Insight Into Hiv-1 Protease Flap Opening. *Journal of Chemical Theory and Computation* 2011,**7**:2348-2352.
9. H. C. Castro, P. A. Abreu, R. B. Geraldo, R. C. A. Martins, R. D. Santos, N. I. V. Loureiro, L. M. Cabral, C. R. Rodrigues. Looking At The Proteases From A Simple Perspective. *Journal of Molecular Recognition* 2011,**24**:165-181.
10. N. Deng, W. Zheng, E. Gallicchio, R. M. Levy. Insights Into The Dynamics Of HIV-1 Protease: A Kinetic Network Model Constructed From Atomistic Simulations. *Journal of the American Chemical Society* 2011,**133**:9387-9394.

11. G. Hu, Q. Zhang, L. Y. Chen. Insights Into scFv: Drug Binding Using The Molecular Dynamics Simulation And Free Energy Calculation. *Journal of Molecular Modeling* 2011,**17**:1919-1926.
12. S. Karthik, S. Senapati. Dynamic Flaps In HIV-1 Protease Adopt Unique Ordering At Different Stages In The Catalytic Cycle. *Proteins: Structure, Function, and Bioinformatics* 2011,**79**:1830-1840.
13. M. E. S. Soliman. Computational Modelling Of Glycosidase Mechanisms: Structural And Mechanistic Aspects [PhD thesis]. United Kingdom: University of Bath; 2011.
14. M. M. Makatini, K. Petzold, S. N. Sriharsha, M. E. S. Soliman, B. Honarparvar, P. I. Arvidsson, Y. Sayed, P. Govender, G. E. M. Maguire , H. G. Kruger, T. Govender. Pentacycloundecane-Based Inhibitors Of Wild-Type C-South African HIV-Protease. *Bioorganic & Medicinal Chemistry Letters* 2011,**21**:2274-2277.
15. M. M. Makatini, K. Petzold , S. N. Sriharsha , N. Ndlovu , M. E.S. Soliman , B. Honarparvar , R. Parboosing , A. Naidoo , P. I. Arvidsson, Y. Sayed , P. Govender , G. E.M. Maguire , H. G. Kruger , T. Govender. 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* 2011,**46**:3976-3986.
16. M. Makatini, T. Chetty, O. K. Onajole, T. Govender, P. Govender, G. E. M. Maguire, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Tetrapeptides. *Journal of Peptide Science* 2012,**18**:121.
17. M. M. Makatini. Design, Synthesis And Screening Of Novel PCU-Peptide/Peptiod Derived HIV Protease Inhibitors [PhD thesis]. Durban: University of KwaZulu-Natal; 2011.
18. O. K. Onajole, Y. Coovadia, T. Govender, H. G.Kruger, G. E. M. Maguire, D. Naidu, N. Singh, P. Govender. In Vitro Antifungal And Antibacterial Activities Of Pentacycloundecane Tetra-Amines. *Chemical Biology and Drug Design* 2011,**77**:295-300.
19. P. Sharma. An Assessment Of The Conformational Profile Of Bombensin And Its Mammalian Analogues Using Computational Chemistry Methods [PhD thesis]. Durban: Durban University of Technology; 2011.

20. Joint United Nations Programme On HIV/AIDS World AIDS Day Report 2011. In: http://www.unaids.org 14 November 2012.
21. L. Menendez-Arias. Targeting HIV: Antiretroviral Therapy And Development Of Drug Resistance. *Trends Pharmacological Science* 2002,**23**:381-388.
22. E. Wood, M. J. Milloy, J. S. G. Montaner. HIV Treatment As Prevention Among Injection Drug Users. *Current Opinoin HIV AIDS* 2012,**7**:151-156.
23. A. R. Zolopa. The Evolution Of HIV Treatment Guidelines: Current State-Of-The-Art Of Art. *Antiviral Research* 2010,**85**:241-244.
24. A. Brik, C. Wong. HIV-1 Protease : Mechanism And Drug Discovery. *Organic and Biomolecular Chemistry* 2003,**1**:5-15.
25. W. R. P. Scott, C. A. Schiffer. Curling Of Flap Tips In HIV-1 Protease As A Mechanism For Substrate Entry And Tolerance Of Drug Resistance. *Structure* 2000,**8**:1259-1265.
26. V. Hornak, A. Okur, R. C. Rizzo, C. Simmerling. HIV-1 Protease Flaps Spontaneously Close To The Correct Structure In Simulations Following Manual Placement Of An Inhibitor Into The Open State. *Journal of the American Chemical Society* 2006,**128**:2812-2813.
27. V. Hornak, A. Okur, R. C. Rizzo, C. Simmerling. HIV-1 Protease Flaps Spontaneously Open And Reclose In Molecular Dynamics Simulations. *Proceedings of the National Academy of Sciences of the United States of America* 2006,**103**:915-920.
28. G. Toth, A. Borics. Closing Of The Flaps Of HIV-1 Protease Induced By Substrate Binding: A Model Of A Flap Closing Mechanism In Retroviral Aspartic Proteases. *Biochemistry* 2006,**45**:6606-6614.
29. G. Toth, A. Borics. Flap Opening Mechanism Of HIV-1 Protease. *Journal of Molecular Graphics and Modelling* 2006,**24**:465-474.
30. V. Hornak, C. Simmerling. Targeting Structural Flexibility In HIV-1 Protease Inhibitor Binding. *Drug Discovery Today* 2007,**12**:132-138.
31. D. I. Freedberg, R. Ishima, J. Jacob, Y. X. Wang, I. Kustanovich, J. M. Louis, D. A. Torchia. Rapid Structural Fluctuations Of The Free HIV Protease Flaps In Solution: Relationship To Crystal Structures And Comparison With Predictions Of Dynamics Calculations. *Protein Science* 2002,**11**:221-232.

32. R. G. Jain, E. S. Furfine, L. Pedneault, A. J. White, J. M. Lenhard. Metabolic Complications Associated With Antiretroviral Therapy. *Antiviral Research* 2001,**51**:151-177.

33. D. Plewczynski, M. Lazniewski, R. Augustyniak, K. Ginalski. Can We Trust Docking Results? Evaluation Of Seven Commonly Used Programs On PDBbind Database. *Journal of Computational Chemistry* 2010,**32**:742-756.

34. S. K. Sadiq, G. De Fabritiis. Explicit Solvent Dynamics And Energetics Of HIV-1 Protease Flap Opening And Closing. *Proteins: Structure, Function, and Bioinformatics* 2010,**78**:2873-2885.

35. D. Li, B. Ji, K. Hwang, Y. Huang. Crucial Roles Of The Subnanosecond Local Dynamics Of The Flap Tips In The Global Conformational Changes Of Hiv-1 Protease. *Journal of Physical Chemistry B* 2010,**114**:3060-3069.

36. T. Lu, Y. Chen, X. Li. An Insight Into The Opening Path To Semi-Open Conformation Of HIV-1 Protease By Molecular Dynamics Simulation. *AIDS* 2010,**24**:1121-1125.

37. M. S. Altaib, P. I. Arvidsson, T. Govender, G. E. M. Maguire, M. Makatini, O. K. Onajolea, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Pentacycloundecane-Based Peptides. *Magnetic Resonance in Chemistry* 2010,**48**:453-461.

38. A. H. Robbins, R. M. Coman, E. Bracho-Sanchez, M. A. Fernandez, C. T. Gilliland, M. Li, M. Agbandje-McKenna, A. Wlodawer, B. M. Dunn, R. McKenna. Structure Of The Unbound Form Of HIV-1 Subtype A Protease: Comparison With Unbound Forms Of Proteases From Other HIV Subtypes. *Acta Crystallographica Section D-Biological Crystallography* 2010,**66**:233-242.

39. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, 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, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, Kollman, P. A. AMBER 11. In. San Francisco: University of California; 2010.

40. R. Karpoormath, P. Govender, H. G. Kruger, T. Govender, G. E. M. Maguire. N,N'-[(8-endo,11-endo-Dihydroxypentacyclo[5.4.0.02,6.03,10.05,9]undecane-8,11-diyl)bis(methylenecarbonyl)]di-L- phenylalanine. *Acta Crystallographica Section E: Structure Reports Online* 2010,**E66**:2.

41. E. Katoh, J. M. Louis, T. Yamazaki, A. M. Gronenborn, D. A. Torchia, R. Ishima. A Solution NMR Study Of The Binding Kinetics And The Internal Dynamics Of An HIV-1 Protease-Substrate Complex. *Protein Science* 2003,**12**:1376-1385.

42. G. Leonis, Z. Czyznikowska, G. Megariotis, H. Reis, M. G. Papadopoulos. Computational Studies Of Darunavir Into HIV-1 Protease And DMPC Bilayer: Necessary Conditions For Effective Binding And The Role Of The Flaps. *Journal of Chemical Information and Modeling* 2012,**52**:1542-1558.

43. A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, S. B. H. Kent. Conserved Folding In Retroviral Proteases - Crystal-Structure Of A Synthetic HIV-1 Protease. *Science* 1989,**245**:616-621.

44. Z. W. Zhu, D. I. Schuster, M. E. Tuckerman. Molecular Dynamics Study Of The Connection Between Flap Closing And Binding Of Fullerene-Based Inhibitors Of The HIV-1 Protease. *Biochemistry* 2003,**42**:1326-1333.

45. H. Heaslet, R. Rosenfeld, M. Giffin, Y. Lin, K. Tam, B. E. Torbett, John H. Elder, D. E. McRee, C. D. Stout. Conformational Flexibility In The Flap Domains Of Ligand-Free HIV Protease. *Acta Crystallographica Section D-Biological Crystallography* 2007,**63**:866-875.

46. L. Galiano, M. Bonora, G. E. Fanucci. Interflap Distances In HIV-1 Protease Determined By Pulsed EPR Measurements. *Journal of the American Chemical Society* 2007,**129**:11004-11005.

47. V. Y. Torbeev, H. Raghuraman, K. Mandal, S. Senapati, E. Perozo, S. B. H. Kent. Dynamics Of "Flap" Structures In Three HIV-1 Protease/Inhibitor Complexes Probed By Total Chemical Synthesis And Pulse-EPR Spectroscopy. *Journal of the American Chemical Society* 2009,**131**:884-885.

48. B. R. Meher, Y. Wang. Interaction Of I50V Mutant And I50L/A71V Double Mutant HIV-Protease With Inhibitor TMC114 (Darunavir): Molecular

Dynamics Simulation And Binding Free Energy Studies. *Journal of Physical Chemistry B* 2012,**116**:1884-1900.

49. S. A. Seibold, R. I. Cukier. A Molecular Dynamics Study Comparing A Wild-Type With A Multiple Drug Resistant HIV Protease: Differences In Flap And Aspartate 25 Cavity Dimensions. *Proteins: Structure, Function, and Bioinformatics* 2007,**69**:551-565.
50. P. Naicker, I. Achilonu, S. Fannuchi, M. Fernandes, M. A. A. Ibrahim, H. W. Dirr, Soliman, M. E. S., Y. Sayed. Structural Insights Into The South African HIV-1 Subtype C Protease: Importance Of Hinge Region Dynamics And Flap Flexibility In Drug Resistance. *Journal of Biomolecular Structure and Dynamics* 2012 Manuscript accepted.
51. B. Honarparvar, M. M. Makatini, K. Petzold, M. E. S. Soliman, P. I. Arvidsson, Y. Sayed, T. Govender, G. E. M. Maguire, H. G. Kruger. Pentacycloundecane-Diol Based HIV-1 Protease Inhibitors: Biological Screening, 2D-NMR And Molecular Simulations Studies. *Chemical Medicinal Chemistry* 2012,**7**:1009-1020.
52. M. Makatini, K. Petzold, P. I. Arvidsson, B. Honarparvar, T. Govender, G. E. M. Maguire, R. Parboosing, Y. Sayed, M. E. S. Soliman, H. G. Kruger. Synthesis, Screening And Computational Investigation Of Pentacycloundecane-Peptoids As Potent CSA-HIV PR Inhibitors. *European Journal of Medicinal Chemistry* 2012 **57**:459-468.
53. M. Makatini, K. Petzold, Alves., C. N., N. Ndlovu, B. Honarparvar, P. Govender, T. Govender, H. G. Kruger, Y. Sayed, J. Lameira, G. E. M. Maguire, Soliman, M. E. S. Synthesis, 2D-NMR And Molecular Modelling Studies Of Pentacyclo-Undecane Lactam Peptides And Peptoids As Potential HIV-1 Wild type C-SA Protease Inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry* 2013,**28**:78-88.
54. R. Karpoormath, Y. Sayed, P. Govender, T. Govender, H. G. Kruger, M. E. S. Soliman, G. E. M. Maguire. 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* 2012,**40**:19-30.

55. D. J. Kempf, K. C. Marsh, J. F. Denissen, E. Mcdonald, S. Vasavanonda, C. A. Flentge, B. E. Green, L. Fino, C. H. Parks, X. Kong, N. E. Wideburg, A. Saldivar, L. Ruizll, W. M. Kati, H. L. Sham, T. Robins, K. D. Stewarti, A. Hsu, J. J. Plattner, J. M. Leonardtt, D. W. Norbeck. ABT-538 Is A Potent Inhibitor Of Human Immunodeficiency Virus Protease And Has High Oral Bioavailability In Humans. *Proceedings of the National Academy of Sciences of the United States of America* 1995,**92**:2488-2493.

56. I. T. Weber, J. Wu, J. Adomat, R. W. Harrison, A. R. Kimmel, E. M. Wondrak, J. M. Louis. Crystallographic Analysis Of Human Immunodeficiency Virus 1 Protease With An Analog Of The Conserved CA-p2 Substrate - Interactions With Frequently Occurring Glutamic Acid Residue At P2' Position Of Substrates. *European Journal of Biochemistry* 1997,**249**:523-530.

57. HyperChem. In. 6.0 ed. 1115 NW 4th street, Gainsville, Florida 32601, USA: Hypercube Inc. pp. HyperChem.

58. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin. UCSF Chimera : A Visualizattion System For Exploratory Research And Analysis. *Journal of Computational Chemistry* 2004,**13**:1605-1612.

59. S. Mosebi, L. Morris, H.W. Dirr, Y. Sayed. Active-Site Mutations In The South African Human Immunodeficiency Virus Type 1 Subtype C Protease Have A Significant Impact On Clinical Inhibitor Binding: Kinetic And Thermodynamic Study. *Journal of Virology* 2008,**82**:11476-11488.

60. S. A. Pawar, A. M. Jabgunde, P. Govender, G. E. M. Maguire, Kruger, H. G., R. Parboosing, M. E. S. Soliman, Y. Sayed, D. D. Dhavale, T. Govender. Synthesis And Molecular Modelling Studies Of Novel Carbapeptide Analogs For Inhibition Of HIV-1 Protease. *European Journal of Medicinal Chemistry* 2012,**53C**:13-21.

61. L. M. F. Gonzalez, R. M. Brindeiro, M. Tarin, A. Calazans, M. A. Soares, S. Cassol, A. Tanuri. In Vitro Hypersusceptibility Of Human Immunodeficiency Virus Type 1 Subtype C Protease To Lopinavir. *Antimicrobial Agents and Chemotherapy* 2003,**47**:2817-2822.

62. W. Humphrey, A. Dalke, K. Schulten. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics* 1996,**14**:33-38.

63. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson. Automated Docking Using A Lamarckian Genetic Algorithm And An Empirical Binding Free Energy Function. *Journal of Computational Chemistry* 1998,**19**:1639-1663.

64. R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell. A Semiempirical Free Energy Force Field With Charge-Based Desolvation. *Journal of Computational Chemistry* 2007,**28**:1145-1153.

65. M. F. Sanner. Python: A Programming Language For Software Integration And Development. *Journal of Molecular Graphics and Modelling* 1999, **17** 57-62.

66. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison. Avogadro: An Advanced Semantic Chemical Editor, Visualization, And Analysis Platform. *Journal of Cheminformatics* 2012,**4**:17.

67. S. M. Ahmed, H. G. Kruger, T. Govender, G. E. M. Maguire, Y. Sayed, M. Ibrahim, P. Naicker, M. E. S. Soliman. Comparison Of The Molecular Dynamics And Calculated Binding Free Energies For Nine FDA-approved HIV-1 PR Drugs Against Subtype B and CSA HIV PR. *Chemical Biology and Drug Design* 2013,**81**:208-218.

68. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. Development And Testing Of A General Amber Force Field. *Journal of Computational Chemistry* 2004,**25**:1157-1174.

69. T. Darden, D. York, L. A. W. Pedersen. Particle Mesh Ewald - An N.Log(N) Method For Ewald Sums In Large Systems. *Journal of Chemical Physics* 1993,**98**:10089-10092.

70. J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen. Numerical-Integration Of Cartesian Equations Of Motion Of A System With Constraints - Molecular-Dynamics Of N-Alkanes. *Journal of Computational Physics* 1977,**23**:327-341.

71. S. Nose. A Molecular-Dynamics Method For Simulations In The Canonical Ensemble. *Molecular Physics* 1984,**52**:255-268.

72. S. Nose. A Unified Formulation Of The Constant Temperature Molecular-Dynamics Methods. *Journal of Chemical Physics* 1984,**81**:511-519.

73. W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson. Semianalytical Treatment Of Solvation For Molecular Mechanics And Dynamics. *Journal of the American Chemical Society* 1990,**112**:6127-6129.

74. T. Hou, R. Yu. 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* 2007,**50**:1177-1188.

Chapter 4

Virtual Screening of PCU-diol peptides as potential C-SA HIV-1 PR

4.1 Abstract

Virtual screening of a total of 640 cage-peptide based inhibitors was performed to gain insight into the structure-activity relationship of these potential inhibitors against C-SA HIV PR. The overall results showed that compounds with the amino terminal (N→C connectivity) had improved binding energies when compared to the carboxyl terminal (C→N connectivity). The most active agents were found to be the amino terminal compounds with a linker distance of n = 2 and 3. There was no real trend in the chirality effect however more prominent energy differences were observed in some diastereomers with the carboxyl and amino terminal for n = 3 and n = 2, respectively. Compounds with bulky hydrophobic amino acids were found to be the most favoured when compared to less hydrophobic, polar, charged, basic and acidic residues in both carboxyl and amino terminal.

4.2 Introduction

The active site pocket of the HIV-1 PR is made up with several well-known subsites where the respective side chains of the substrates are positioned upon binding[1]. It has been reported in literature that a phenylalanine residue frequently occupies the S_1 subsite and 40% of the *Gag-pol* polyprotein precursor gene sequences of the virus consist of phenylalanine[2, 3]. It is also known from literature that the S_1/S_1' subsites preferentially accommodates bulky hydrophobic residues (such as phenylalanine) while the S_2/S_2' and S_3/S_3' subsites prefer both hydrophilic and hydrophobic residues such as valine and alanine[4, 5].

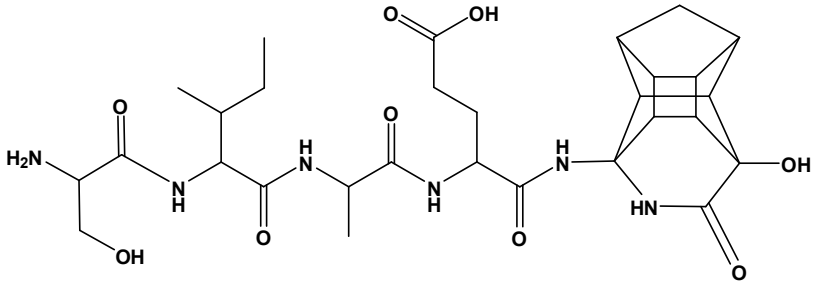
There are currently eleven commercial HIV-1 PR inhibitors approved by the Food and Drug Administration (FDA) in the USA and a number of others are currently undergoing clinical trials[6].

Previous non-HIV related studies reported that the inclusion of a hydrocarbon cage skeleton into biologically active compounds advanced their function as a result of the ability of the drug to cross the blood-brain barrier and to penetrate the central nervous system due to the hydrophobicity of the

cage. It also reduces the metabolic degradation of the drug thereby prolonging its pharmaceutical effect in the body[7-11].

We have reported the synthesis[5, 11-13], computational studies[3, 5, 9, 14, 15], NMR [10, 12, 13] and biological testing[3, 5] of a series of pentacycloundecane (PCU)-based peptides and peptoids as potential South African HIV-1 protease subtype (C-SA HIV PR) inhibitors. The best inhibitory activities (IC_{50}) observed from *in vitro* analysis for both cage lactam peptides (Figure 41, represent compound **1** with 78 nM) and peptoids in the 70 nM range. It was concluded from these comprehensive studies, that the PCU skeleton occupies the S_1/S_1' subsites while the peptide side chains fit the S_2/S_2' and S_3/S_3' sites.

Moreover, these studies confirmed that variation of the amino acids of the peptide side chain (or the corresponding peptoids) has a remarkable influence on the inhibitory activity of these compounds towards the C-SA HIV PR. We have previously demonstrated that the chirality of the cage is important and as expected, leading to different observed HIV inhibition activities[5].



1 (IC_{50} = 78 nM)

Figure 41. Structures of the previously synthesized and biologically tested PCU-lactam peptide[5].

Our previous studies concluded, that the docking results for these compounds give reasonable agreement with the experimentally observed inhibition activities (IC_{50})[16]. However, there are two potential reasons for not getting a correlation between the docked results and IC_{50}: (a) the docking binding energies are not sophisticated enough and (b) as we have shown[17], experimental IC_{50} results tend to vary from one reported literature value to the next.

It appeared to us that the symmetrical cage diol compounds[18] which have two peptide side chains (**2a** and **2b,** Figure 42) should in theory be able to form more hydrogen bonds with the PR active site than the cage lactam with one peptide side chain (**1**) only (see Figure 41). It is generally

accepted that inhibitors, which form more hydrogen bonds with the active site, exhibit better activities[19-21].

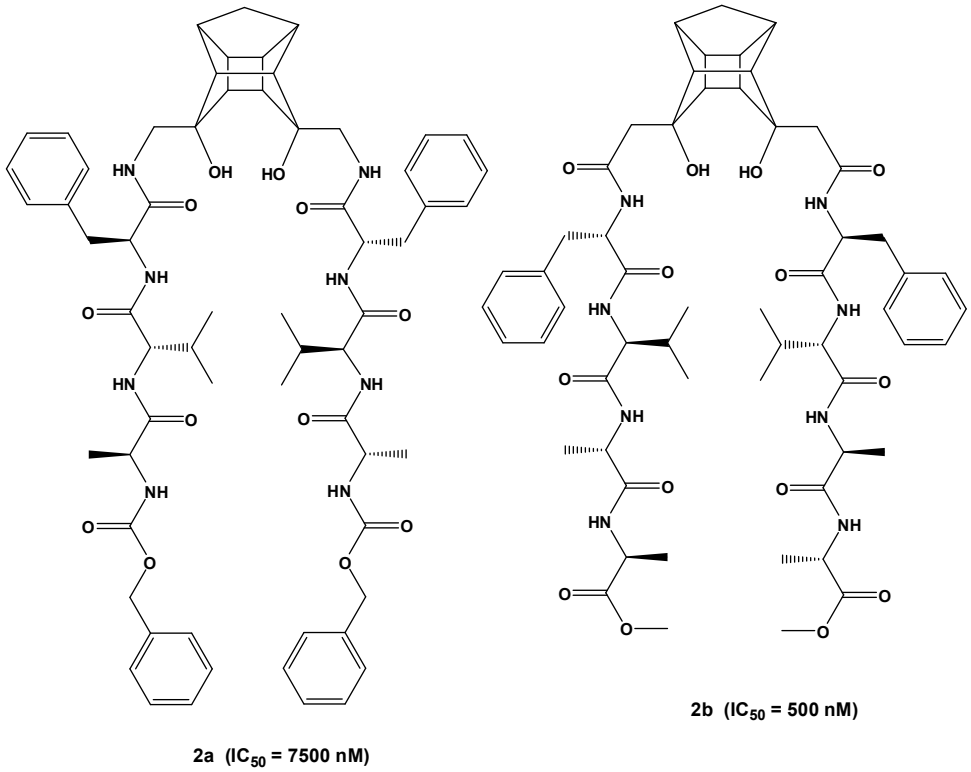


Figure 42. Structures of the previously synthesized and biologically tested PCU-diol peptides against C-SA HIV-1 PR[18].

In general, these cage diol peptides exhibited more than an order of magnitude lower activity against HIV PR than the best lactam peptide or peptoid.

It was argued[18] that variation of the length of the linker[22-25] between the cage and the peptide combined with a variation of the amino acid chirality on either side, may enable us to design more active cage diol peptides (see Figure 43).

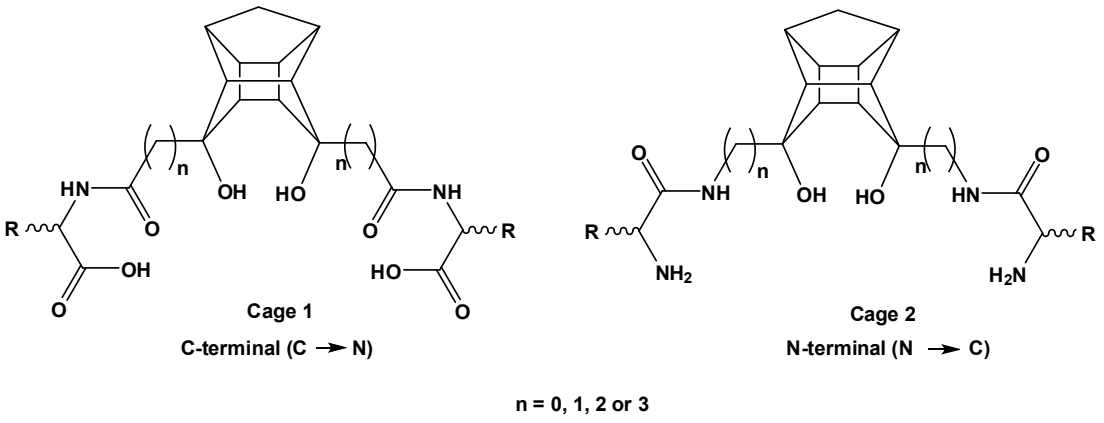


Figure 43. Model structure for the PCU-diol peptides to be investigated. The R groups are represented by the nineteen natural amino acids as well as 2-phenylglycine (Phg) with four different diastereomers for both cage**1** and cage**2** (20 x 4 x 2 = 160 compounds). This was done for n = 0 – 3 (160 x 4 = 640 compounds).

It is important, to note that some of these compounds (such as cage**2**, n = 0, 1) would be difficult if not impossible, to synthesise.

The aim of this study is to propose and test a suitable computational model to further optimize the potential of the PCU-diol compounds against C-SA HIV-1 PR, (see Figure 43, cage**1** and cage**2**). In order to achieve that molecular docking will be used to determine the interactions (binding free energies) of the PCU systems with C-SA PR. With respect to complexation to the PR we will attempt to answer the following questions:

1. To determine whether C→N (R-cage**1**-R) or N→C (R-cage**2**-R) amino acid attachment to the cage gives optimum binding to the PR active site.
2. The optimal linker distance (n = 0, 1, 2 or 3) of these amino acids to the cage skeleton.
3. The most favourable amino acid(s) and its chirality to be attached to the cage-diol skeleton for the best complimentary interaction with HIV-PR (C-SA).

To accomplish these objectives, we conducted an extensive virtual screening study to obtain a theoretical Structure-Activity Relationship (SAR) of a large number of these compounds. It is important to note that the symmetric cage diol becomes chiral (10 of the 11 carbon atoms) when a chiral residue is attached to any (or both) of the two side chains[18]. Variation of the chirality of the amino acid side chains (attached to both sides) therefore, results in the formation of four different diastereomers [(*S,S*), (*R,R*), (*R,S*) and (*S,R*)]. We chose to study one kind of amino acid

attached to both side of the cage because an additional (second) amino acid will protrude out of the enzyme and exhibit limited hydrogen bond contribution.

Docking simulations for all natural amino acid and 2-phenylglycine with four different diastereomers at different separation distances (n) from the cage framework were performed for both cage**1** and cage**2** (Figure 43). This involved 640 different molecules.

4.3 Materials and methods

4.3.1 Computational simulations

The following sections present an outline about the preparation of the ligands and receptors for molecular modelling. It also provides details of the software and the parameters that were employed.

4.3.1.1 Preparation of PCU-derived peptide inhibitors

PCU-diol peptide-based compounds were constructed and optimized with the molecular mechanics optimization force field implemented in Hyperchem 6.0 software[26] and the dihedral angles for all amide bonds were 180°. The *trans* conformation of the amide bond requires 88 kJ/mol (21 kcal/mol) for rotation about C-N bond to form the cis conformation[27]. Since this does not occur at room temperature, the *trans* amide bond was fixed during docking. Nineteen natural amino acids as well as 2-phenylglycine (Phg) with four different diastereomers were studied for both cage**1** and cage**2** (20 x 4 x 2 = 160 compounds). This was done for n = 0 – 3 (160 x 4 = 640 compounds). The ionization states of the aliphatic amine and carboxylic acids groups of the PCU-based peptide inhibitors were reported to have no significant effects on the docking results when they are either in the neutral or ionized states[15]. These resulted in the screening of a total of 640 neutral compounds against C-SA HIV-1 protease. The minimized structures were then subjected to docking studies.

4.3.1.2 C-SA HIV-1 PR enzyme model system

The model structure of the C-SA HIV-1 PR has been explained in detail in Section 3.3.1.2

4.3.1.3 Docking of the inhibitors into the active site of the C-SA HIV-1 PR model

Docking procedure of the 640 compounds against C-SA HIV-1 PR was performed as in Section 3.3.1.3.

4.4 Results and discussion

Virtual screening

In all the investigated amino acids with n = 0 of cage**1** compounds (see Figure 44), the *(S)*-Trp-cage**1**-Trp-*(R)* was observed to be the best with a figure of -9.29 kcal/mol followed by *(R)*-Pro-cage**1**-Pro-*(R)*, *(S)*-Phg-cage**1**-Phg-*(R)*, *(R)*-Tyr-cage**1**-Try-*(S)* and *(R)*-Phe-cage**1**-Phe-*(R)* with -8.45 kcal/mol, -8.39 kcal/mol, -8.37 kcal/mol and -7.86 kcal/mol, respectively. When looking at the chirality effect for the most favoured amino acids with strong binding, it was found that there is no real trend in the binding energies.

Thus, the attachment of bulky hydrophobic residues (Trp, Pro, Phg, Tyr and Phe) for some diastereomers directly to the PCU cage results in the most favourable binding energies with the PR in comparison to the less hydrophobic (Ala, Ile, Leu and Val), polar (Asn, Met, Cys, Ser, Gln and Thr), charged basic residues (Arg, His and Lys). The poorest results were obtained with the acidic residues Asp and Glu. The bulky hydrophobic residues are expected to occupy S_2/S_2'subsites since the cage skeleton occupies the S_1/S_1'subsites.

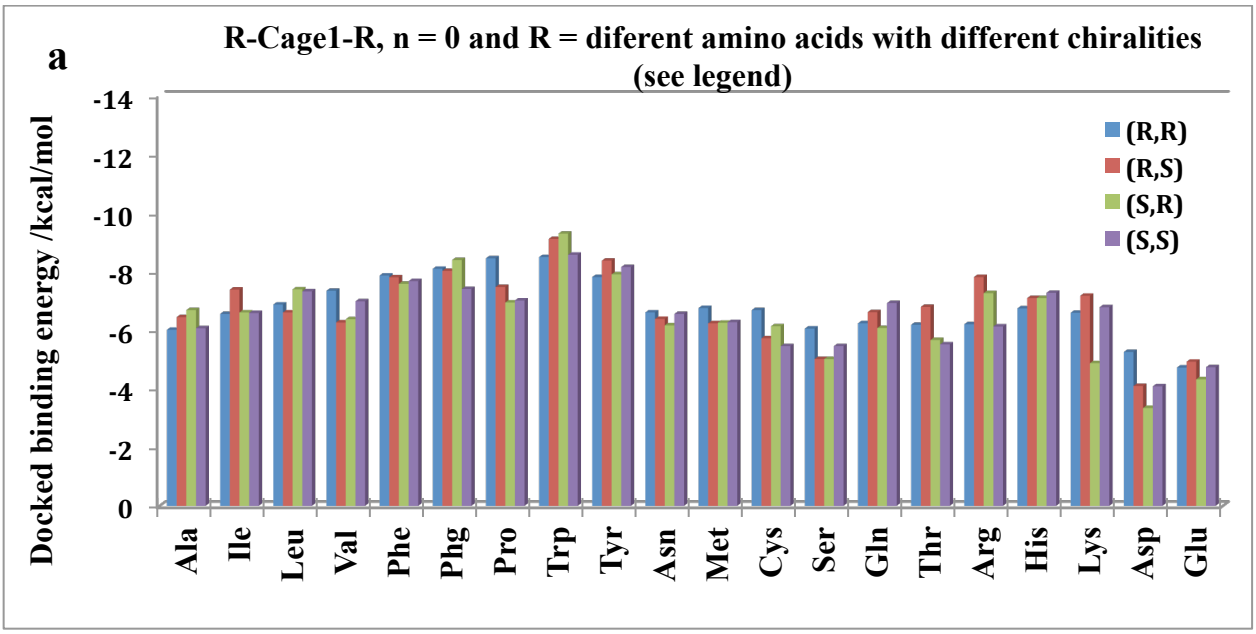


Figure 44. A graph representing the virtual screening results of all 20 investigated amino acids for the carboxyl terminal, when n = 0 for R-cage**1**-R.

The best result in Figure 45, when n = 1 for cage**1**, was observed for *(S)*-Trp-cage**1**-Trp-*(R)* with -8.69 kcal/mol as was the case in Figure 44. Other compounds, such as *(S)*-Phe-cage**1**-Phe-*(R)*, *(R)*-

Tyr-cage**1**-Tyr-*(R)*, *(R)*-Phg-cage**1**-Phg-*(S)* and *(S)*-Ile-cage**1**-Ile-*(S)* however showed better binding with binding energies of -8.51 kcal/mol, -8.46 kcal/mol, -8.23 kcal/mol and -7.91 kcal/mol, respectively but with different trends to Figure 44. Similarly, it appears that there is no obvious trend in the binding energies with relation to any particular chiral arrangement.

The *(S)*-Trp-cage**1**-Trp-*(R)* inhibitor showed the binding energy of -9.29 kcal/mol in Figure 44 while in Figure 45 it was -8.69 kcal/mol. This indicates that increasing the linker from n = 0 to n = 1 resulted in a decrease in the binding activity. A decrease in the binding energies were also observed for the *(R)*-Pro-cage**1**-Pro-*(R)*, *(S)*-Phg-cage**1**-Phg-*(R)* and *(R)*-Tyr-cage**1**-Tyr-*(S)* inhibitors. The binding energies for these compounds in Figure 44 were -8.45 kcal/mol, -8.39 kcal/mol and -8.37 kcal/mol while in Figure 45 they were -6.81 kcal/mol, -7.90 kcal/mol and -7.29 kcal/mol, respectively. In contrast the *(R)*-Phe-cage**1**-Phe-*(R)* inhibitor showed an increase in the binding energy from n = 0 to n = 1. When n = 0 the binding energy was -7.86 kcal/mol and when n = 1 it was -8.30 kcal/mol.

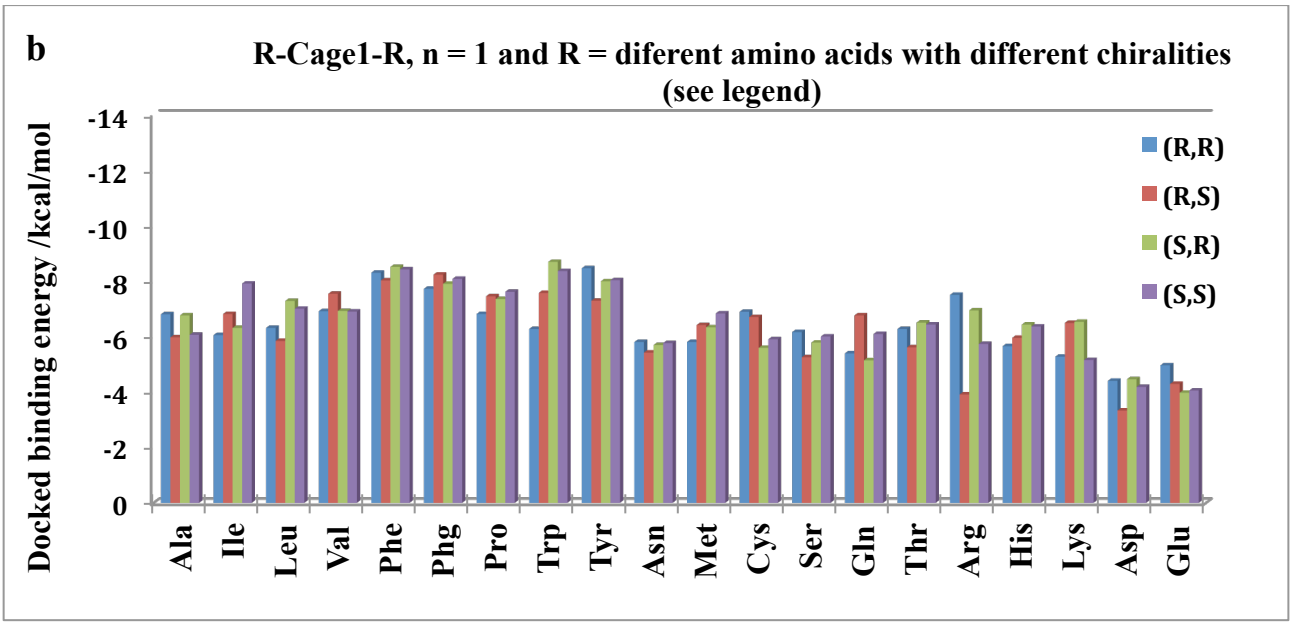


Figure 45. A graph representing the virtual screening results of all 20 investigated amino acids for the carboxyl terminal, when n = 1 for R-cage**1**-R.

As the number of methylene spacers increases it is believed that the amino acids residues will occupy other subsites. The best inhibitors for cage**1** when n = 2, (see Figure 46) were *(R)*-Phe-cage**1**-Phe-*(S)*, *(S)*-Trp-cage**1**-Trp-*(S)*, *(S)*-Phg-cage**1**-Phg-*(S)*, *(S)*-Tyr-cage**1**-Tyr-*(R)* and *(S)*-Leu-

cage**1**-Leu-*(S)* with binding energies of -8.87 kcal/mol, -8.81 kcal/mol, -8.61 kcal/mol, -8.46 kcal/mol and -8.26 kcal/mol, respectively. There was a significant increase in the binding ability for the *(S)*-Leu-cage**1**-Leu-*(S)*.

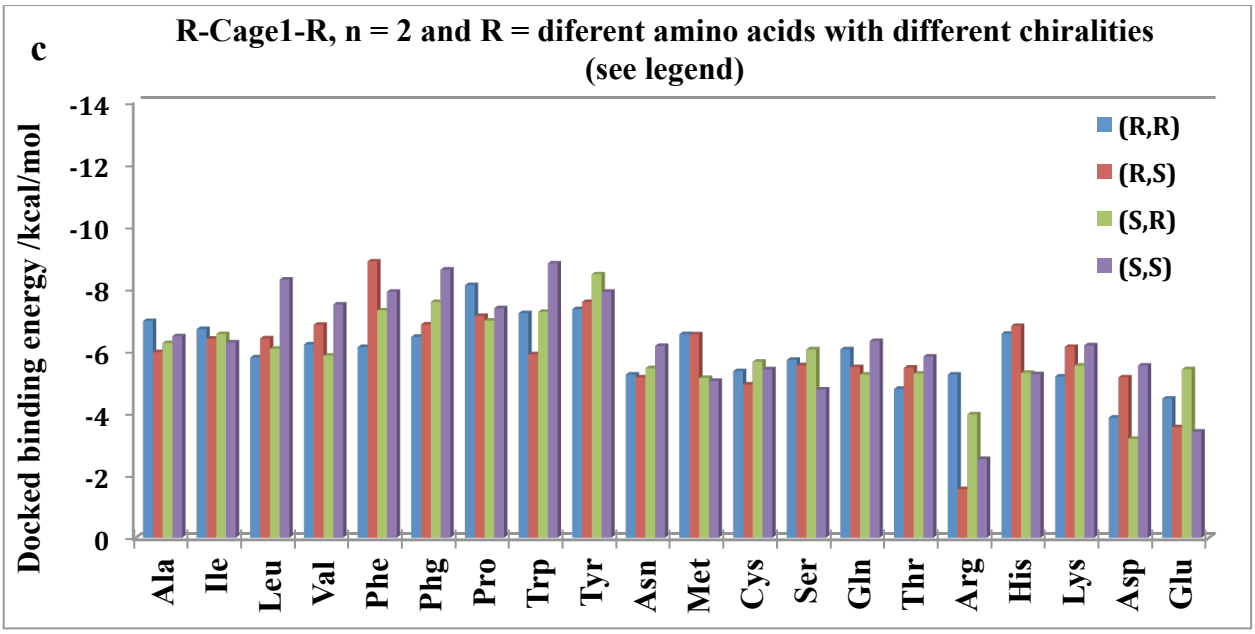


Figure 46. A graph representing the virtual screening results of all 20 investigated amino acids for the carboxyl terminal, when n = 2 for R-cage**1**-R.

As the distance of the linker between the cage skeleton and the side chains increased further (n = 3), the *(R)*-Phe-cage**1**-Phe-*(R)* was the best inhibitor with -9.94 kcal/mol, (see Figure 47). Other bulky compounds, such as *(S)*-Phg-cage**1**-Phg-*(R), (S)*-Pro-cage**1**-Pro-*(R)* and *(R)*-Trp-cage**1**-Trp-*(S)* gave the most promising complexing energies with -8.80 kcal/mol, -7.95 kcal/mol, -7.77 kcal/mol, respectively. For these compounds once again, there was no real trend in the chirality effect. Interestingly, amongst the hydrophilic species there appears to be an increasing trend in favour of one or more particular diastereomer. For example, *(R)*-Lys-cage**1**-Lys-*(R)* renders a uniquely lower energy of -7.41 kcal/mol compared to its other chiral equivalents at around 4 kcal/mol.

The overall results for the cage**1** compounds, with n = 0 – 3, showed that irrespective of the spacer length the bulky hydrophobic amino acids are the best compounds when compared to the less hydrophobic, polar, charged, acidic or basic examples. The investigation of the chirality effect showed different results for each amino acid making it difficult to find a trend. It seems clear that

compounds with Asp and Glu residues would be poor candidates for inhibitor synthesis as they consistently yielded poor results.

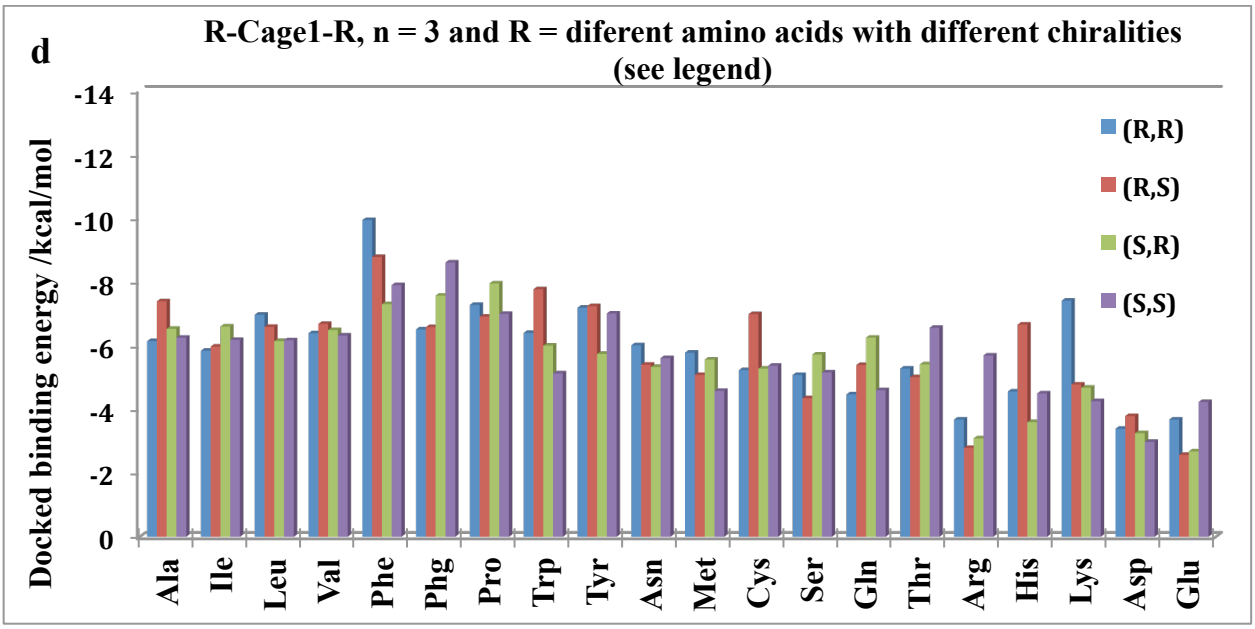


Figure 47. A graph representing the virtual screening results of all 20 investigated amino acids for the carboxyl terminal, when n = 3 for R-cage**1**-R.

Figure 48, shows the cage**2** compounds when n = 0. It was observed that the best inhibitor was (*R*)-Phg-cage**2**-Phg-(*R*) with the binding energy of -11.92 kcal/mol. The other promising compounds were (*R*)-Phe-cage**2**-Phe(*S*), (*R*)-Trp-cage**2**-Trp-(*R*), (*R*)-Tyr-cage**2**-Tyr-(*R*) and (*R*)-Thr-cage**2**-Thr-(*S*) with the binding energies of -11.36 kcal/mol, -11.08 kcal/mol, -10,56 kcal/mol, -9.84 kcal/mol, respectively. The overall energy results in this case for the more hydrophobic amino acids are closer in value to the compounds with bulky hydrophobic residues. It is observed that the inhibitors with N→C connectivity (R-cage**2**-R), exhibit slightly improved binding energies compared to those with C→N examples. These results correlate with the recently published work on PCU-diol peptides.[18]

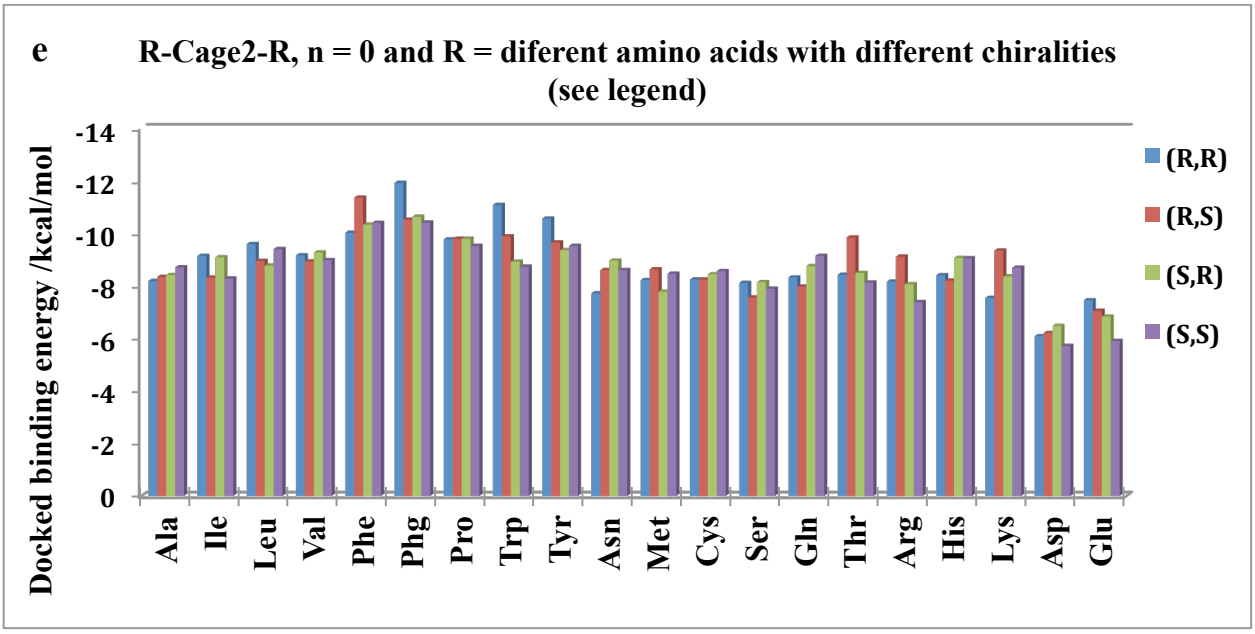


Figure 48. A graph representing the virtual screening results of all 20 investigated amino acids for the amino terminal, when n = 0 for R-cage**2**-R.

When the linker was n = 1, (see Figure 49) in R-cage**2**-R, the best inhibitor was *(S)*-Phg-cage**2**-Phg-*(S)* with -12.22 kcal/mol followed by *(R)*-Phe-cage**2**-Phe-*(R), (R)*-Trp-cage**2**-Trp-*(R), (R)*-Tyr-cage**2**-Tyr-*(R)* with binding energies of -11.41 kcal/mol, -10.83 kcal/mol and -10.44 kcal/mol, respectively. The result for *(S)*-Phg-cage**2**-Phg-*(S)* is the same as for n = 0, but in that case the chirality was reversed.

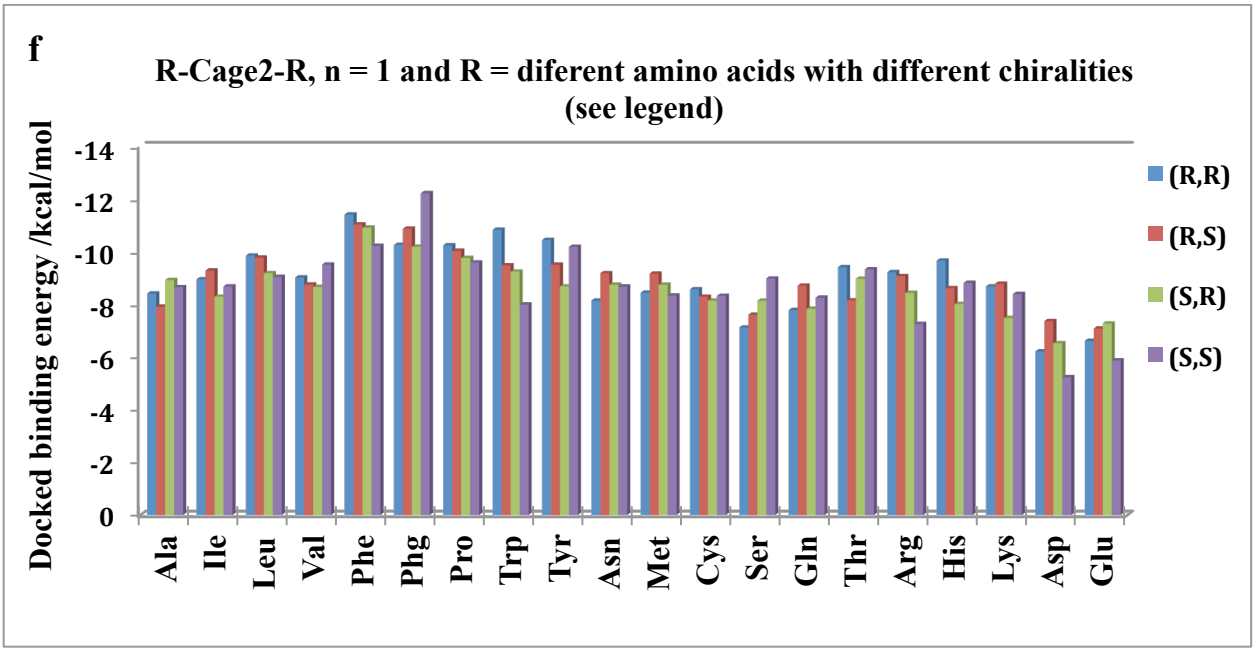


Figure 49. A graph representing the virtual screening results of all 20 investigated amino acids for the amino terminal, when n = 1 for R-cage**2**-R.

For the n = 2 for cage**2**, the best binding energies observed were for *(S)*-Trp-cage**2**-Trp-*(S), (R)*-Phg-cage**2**-Phg-*(R), (S)*-Val-cage**2**-Val-*(R)*, *(S)*-Ile-cage**2**-Ile-*(R)* and *(S)*-Gln-cage**2**-Gln-*(S)* with binding energies of -12.44 kcal/mol, -12.32 kcal/mol, -11.78 kcal/mol, -11.67 kcal/mol and -10.60 kcal/mol, respectively. It is observed that across the graph all residues exhibit improved binding for the amino terminal (N→C connectivity) when n = 2 for cage**2** compounds compared to n = 1 (see Figure 50). Interestingly, a similar trend as was the case for cage**1** n = 3 is observed in these results. In particular there is an increasing trend in favour of one or more particular diastereomer binding energy. In this example however the trend appears across the range of amino acid side chains.

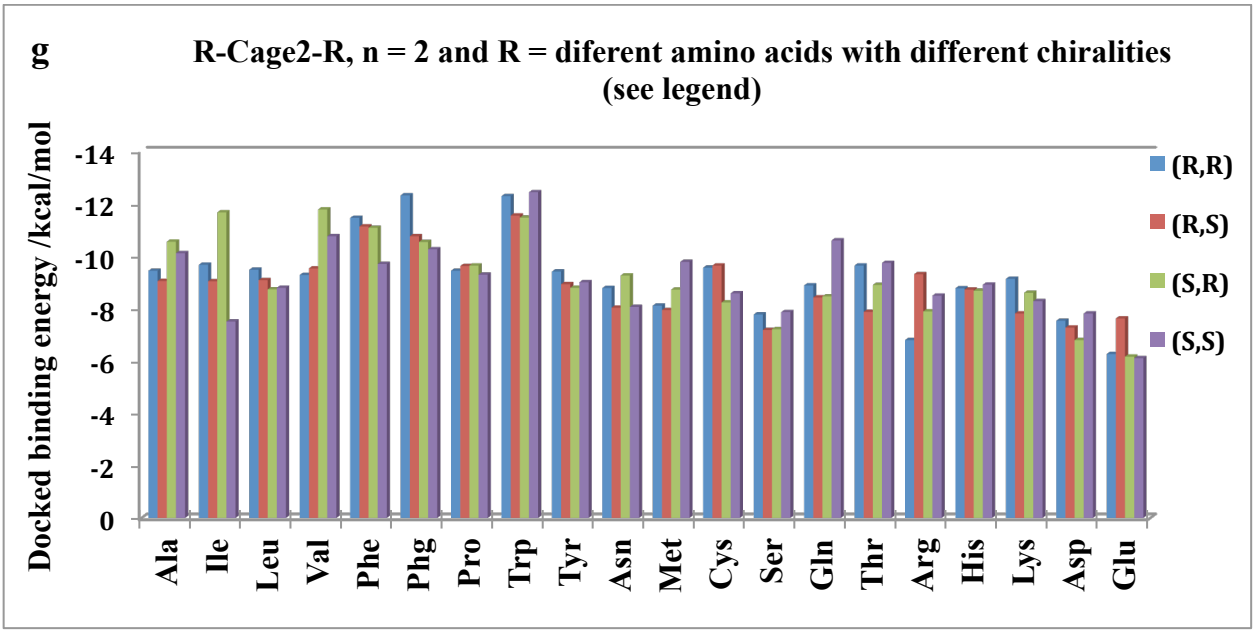


Figure 50. A graph representing the virtual screening results of all 20 investigated amino acids for the amino terminal, when n = 2 for R-cage**2**-R.

When n = 3 (see Figure 51) it was observed that the best compounds were the *(S)*-Phg-cage**2**-Phg-*(R), (S)*-Phe-cage**2**-Phe-*(R), (R)*-Tyr-cage**2**-Tyr-*(R), (R)*-His-cage**2**-His-*(S)* and *(S)*-Val-cage**2**-Val-*(R)* with the binding energies of -12.05 kcal/mol, -11.85 kcal/mol, -10.19 kcal/mol, -9.94 kcal/mol and -9.74 kcal/mol, respectively. The diastereomer binding energy trend observed for n = 2 for cage**2** is lost in these results and the energies appear to become more homogenous.

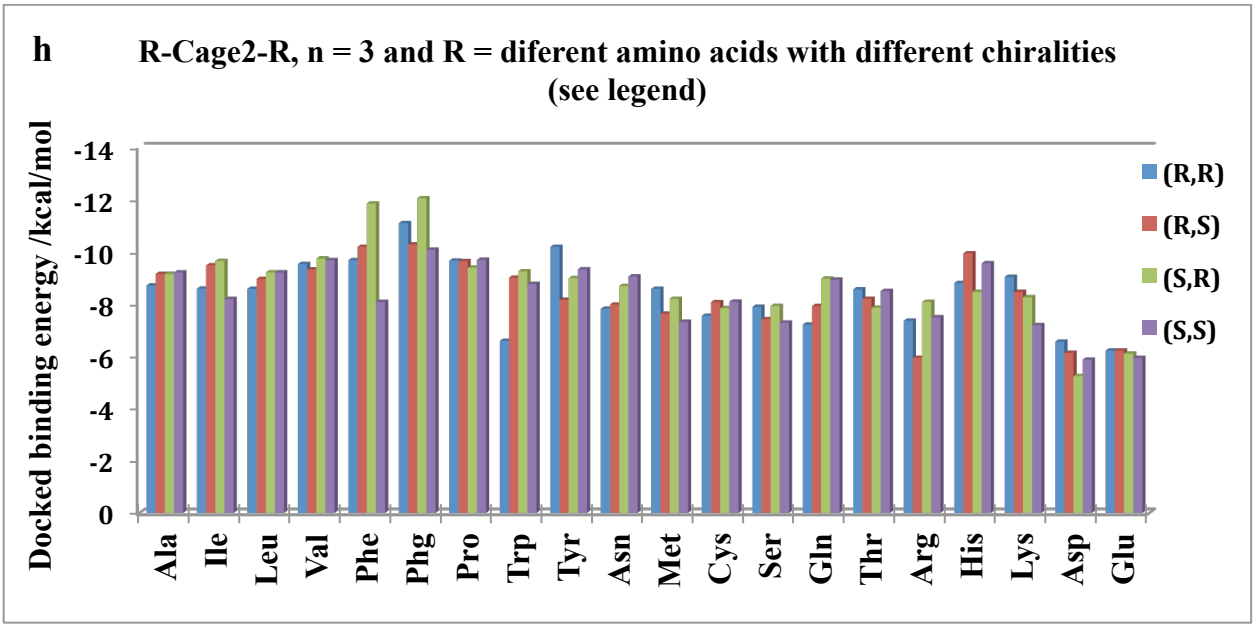


Figure 51. A graph representing the virtual screening results of all 20 investigated amino acids for the amino terminal, when n = 3 for R-cage2-R.

4.5 Conclusion

In the overall results for the carboxyl terminal (C→N connectivity), it was revealed the bulky hydrophobic residues gives the best binding when n = 0 and n = 1, with slightly weaker binding when n = 2 and n = 3. The inhibitors containing acid residues (aspartic acid and glutamic acid) consistently exhibited weak binding for n = 0, 1, 2 and 3. Although there was no real trend in the chirality for the cage**1** compounds when n = 3, some diastereomers gave more prominent energy differences.

The inhibitors with the amino terminal (N→C connectivity) showed improved binding energies when compared with the carboxyl terminal. The most favoured amino acids, once again, were the ones with bulky hydrophobic residues. In this family of compounds, as the spacer length increased the amino acids with more polar residues generally, showed an increase in binding energies. Again, there was no conclusive trend in the chirality for the cage**2** compounds but when n = 2, a number of diastereomers gave more prominent energy differences. Interestingly, in the case of the single "arm" lactam compound **1**, the most effective agent had a hydrophilic amino acid attached to the cage moiety.

As was stated earlier, the synthesis of compounds based on cage**2 (**n = 0,1) would be not be facile. Fortunately, the results of this study demonstrate that the most active compounds would in fact be based on cage**2** where n = 2, 3.

In this study, PCU-diol peptides with ≥ 12 rotatable bonds were investigated. Thus, the docked binding energies are anticipated to be of a crude nature. The study does, however, show promising results for both families of compounds. It is clear that the cage**2** compounds render a larger range of synthetic options as the structure-activity trend is broader. To verify these results, subsequent MD binding free energies calculations need to be undertaken. The final results will then have to be compared with reliable experimental *in vitro* activities.

4.6 References

1. A. Brik, C. Wong. HIV-1 Protease : Mechanism And Drug Discovery. *Organic and Biomolecular Chemistry* 2003,**1**:5-15.
2. J. C. Clemente, R. M. Coman, M. M. Thiaville, L. K. Janka, J. A. Jeung, S. Nukoolkarn, L. Govindasamy, M. Agbandje-McKenna, R. McKenna, W. Leelamanit, M. M. Goodenow, B. M. Dunn. Analysis Of HIV-1 CRF_01 A/E Protease Inhibitor Resistance: Structural Determinants For Maintaining Sensitivity And Developing Resistance To Atazanavir. *Biochemistry* 2006,**45**:5468-5477.
3. R. Karpoormath, Y. Sayed, P. Govender, T. Govender, H. G. Kruger, M. E. S. Soliman, G. E. M. Maguire. 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* 2012,**40**:19-30.
4. A. Wlodawer, J. Vondrasek. Inhibitors Of HIV-1 Protease: A Major Success Of Structure-Assisted Drug Design. *Annual Review of Biophysics and Biomolecular Structure* 1998,**27**:249-286.
5. M. M. Makatini, K. Petzold, S. N. Sriharsha, M. E. S. Soliman, B. Honarparvar, P. I. Arvidsson, Y. Sayed, P. Govender, G. E. M. Maguire , H. G. Kruger, T. Govender. Pentacycloundecane-Based Inhibitors Of Wild-Type C-South African HIV-Protease. *Bioorganic & Medicinal Chemistry Letters* 2011,**21**:2274-2277.
6. FDA. Antiretroviral Drugs Used In The Treatment Of HIV Infection. In: http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm; 27 June 2013.
7. A. P. Marchand. Synthesis And Chemistry Of Homocubanes, Bishomocubanes And Trishomocubanes. *Chemical Reviews* 1989,**89**:1011-1033.
8. G. W. Griffin, A. P. Marchand. Synthesis And Chemistry Of Cubanes. *Chemical Reviews* 1989,**89**:997-1010.
9. K. Bisetty, F. J. Corcho, J. Canto, H. G. Kruger, J. J. Perez. A Theoretical Study Of Pentacyclo-Undecane Cage Peptides Of The Type [Ac-X-Y-NHMe]. *Journal of Peptide Science* 2006,**12**:92-105.
10. M. S. Altaib, P. I. Arvidsson, T. Govender, G. E. M. Maguire, M. Makatini, O. K. Onajolea, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Pentacycloundecane-Based Peptides. *Magnetic Resonance in Chemistry* 2010,**48**:453-461.

11. O. K. Onajole, Y. Coovadia, T. Govender, H. G.Kruger, G. E. M. Maguire, D. Naidu, N. Singh, P. Govender. In Vitro Antifungal And Antibacterial Activities Of Pentacycloundecane Tetra-Amines. *Chemical Biology and Drug Design* 2011,**77**:295-300.
12. O. K. Onajole, T. Govender, M. Makatinia, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Penta-Cycloundecane Amine Derivatives As Potential Antituberculosis Agents. *Magnetic Resonance in Chemistry* 2008,**46**:1007-1014.
13. M. Makatini, T. Chetty, O. K. Onajole, T. Govender, P. Govender, G. E. M. Maguire, H. G. Kruger. Synthesis And NMR Elucidation Of Novel Tetrapeptides. *Journal of Peptide Science* 2012,**18**:121.
14. K. Bisetty, F. J. Corcho, J. Canto, H. G. Kruger, J. J. Perez. Simulated Anneling Study Of The Pentacyclo-Undecane Cage Amino Acid Tripeptides Of The Type [AC-X-Y-Z-NHMe]. *Journal of Molecular Structure: Theochem* 2006,**759**:145-158.
15. M. M. Makatini, K. Petzold , S. N. Sriharsha , N. Ndlovu , M. E.S. Soliman , B. Honarparvar , R. Parboosing , A. Naidoo , P. I. Arvidsson, Y. Sayed , P. Govender , G. E.M. Maguire , H. G. Kruger , T. Govender. 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* 2011,**46**:3976-3986.
16. R. Karpoormath, Y. Sayed, T. Govender, H. G. Kruger, M. E. S. Soliman, G. E. M. Maguire. Novel PCU Cage Diol Peptides As Potential Targets Against Wild-Type CSA HIV-1 Protease: Synthesis, Biological Screening And Molecular Modelling Studies. *Medicinal Chemistry Research* 2013 Accepted.
17. S. M. Ahmed, H. G. Kruger, T. Govender, G. E. M. Maguire, Y. Sayed, M. Ibrahim, P. Naicker, M. E. S. Soliman. Comparison Of The Molecular Dynamics And Calculated Binding Free Energies For Nine FDA-approved HIV-1 PR Drugs Against Subtype B and CSA HIV PR. *Chemical Biology and Drug Design* 2013,**81**:208-218.
18. B. Honarparvar, M. M. Makatini, K. Petzold, M. E. S. Soliman, P. I. Arvidsson, Y. Sayed, T. Govender, G. E. M. Maguire, H. G. Kruger. Pentacycloundecane-Diol Based HIV-1 Protease Inhibitors: Biological Screening, 2D-NMR And Molecular Simulations Studies. *Chemical Medicinal Chemistry* 2012,**7**:1009-1020.
19. A. Velazquez-Campoy, Y. Kiso, E. Freire. The Binding Energetics Of First- And Second-Generation HIV-1 Protease Inhibitors: Implications For Drug Design. *Archives of Biochemistry and Biophysics* 2001,**390**:169-175.
20. N. M. King, M. Prabu-Jeyabalan, E. A. Nalivaika, P. Wigerinck, M. P. de Bethune, C. A. Schiffer. Structural And Thermodynamic Basis For The Binding Of TMC114, A Next-

Generation Human Immunodeficiency Virus Type 1 Protease Inhibitor. *Journal of Virology* 2004,**78**:12012-12021.

21. C. McCoy. Darunavir: A Nonpeptidic Antiretroviral Protease Inhibitor. *Clinical Therapeutics* 2007,**29**:1559-1576.
22. T. G. Levitskaia, B. A. Moyer, P. V. Bonnesen, A. P. Marchand, K. Krishnudu, Z. B. Chen, Z. L. Huang, H. G. Kruger, A. S. McKim. Novel Approach To Sodium Hydroxide Separation: Synergistic Pseudo-Hydroxide Extraction By A Fluorinated Alcohol And Cage-Functionalized Crown Ethers. *Journal of the American Chemical Society* 2001,**123**:12099-12100.
23. G. A. Boyle, T. Govender, R. Karpoormath, H. G. Kruger. exo-8,exo-11-Divinylpentacyclo 5.4.0.0(2,6).0(3,10).0(5,6) undecane-endo -8, endo-11-diol. *Acta Crystallographica Section E: Structure Reports Online* 2007,**63**:O3977-U2352.
24. G. A. Boyle, T. Govender, H. G. Kruger, G. E. M. Maguire. Synthesis Of Chiral Pentacyclo-Undecane Ligands And Their Use In The Enantioselective Alkylation Of Benzaldehyde With Diethylzinc. *Tetrahedron-Asymmetry* 2004,**15**:2661-2666.
25. G. A. Boyle, T. Govender, R. Karpoormath, H. G. Kruger. exo-8,exo-11-Diallylpentacyclo-5.4.0.0(2,6).0(3,10).0(5,9) undecane-end o-8,endo-11-diol. *Acta Crystallographica Section E: Structure Reports Online* 2007,**63**:O4797-U5875.
26. HyperChem. In. 6.0 ed. 1115 NW 4th street, Gainsville, Florida 32601, USA: Hypercube Inc. pp. HyperChem.
27. J. Clayden, N. Greeves, S. Warren, P. Worthers. *Organic Chemistry*: Oxford University Press: New York; 2001.

Chapter 5

Summary

In Chapter 3, molecular dynamics simulation in implicit solvation was performed for the subtype B PR, C-SA PR, PCU-based compounds bound to C-SA PR and ritonavir, natural substrate bound to both proteases (subtype B and C-SA PR). Ritonavir and a natural substrate were used as references.

A comparative study between unbound subtype B and C-SA PR showed that C-SA PR exhibits higher flap flexibility than subtype B PR. From the MD trajectories, it was observed that the subtype B PR opens at 3.4 ns and remained open for the entire simulation while in C-SA PR three conformations were observed *i.e.,* closed, semi-opened and opened forms. The unbound C-SA PR was observed to open at 3.5 ns and then returned to the semi-opened form. A semi-opened form of the protease was observed after the curling conformation (at 2.0 ns), which resulted in an opened conformation at 3.5 ns. The flaps returned to the semi-opened conformation at 4.3 ns. At 5.2 ns, a second opened form was observed and a fully-opened form at 8.6 ns was reached with a flap separation (the Ile149-Ile50 distance) of 28 Å.

When evaluating the extent of flap opening for both complexes and unbound PRs, it was found that all complexes had reduced the normal movement of the PR (subtype B and C-SA PR). When the metric parameter for measuring the distance between the flap tips (Cα Ile50 – Cα Ile149) was used, it was found that the Cα Ile50 – Cα Ile149 metric was affected by flap-curling and flap asymmetry. Measuring the distance between the Cα K55 – Cα K154 seemed to give more accurate results about the extent of flap opening since it was not affected by flap-curling. In addition, it was found that measuring the distance between the flap tips and the catalytic aspartic residues (I50 Cα- Asp25 Cβ) and (I149 Cα- Asp124 Cβ) gives more information about how tightly the inhibitor is held in the pocket. Also, measuring the dihedral angles [CAs of G48-G49-I50 in monomer A (G147-G148-I149 in monomer B)] of the flap tips only shows which angles represent which conformations but not the extent of flap movement.

When looking at RMSF plots, it was found that the Asp25 and Asp124 for both unbound PRs and complexes gave similar distances. The PCU-ether peptide and PCU-lactam peptide showed reduced flap movement while the PCU-lactam peptoid showed no reduction in the flap motion. The PCU-diol peptide and PCU-diol peptoid showed relatively similar reduction of the flap movement. From these observations, it was clear that the PCU based compounds do affect the flap movement but to a lesser degree than ritonavir and the natural substrate.

The PCU-based compounds, have been previously biologically tested but when looking at the MD results and the flap movement measurements, it is clear that a quantitative relationship between the reported activities and the observed flap movements is not identifiable.

In Chapter 4, virtual screening of 640 inhibitors (see Figure 52) was performed with AutoDock.

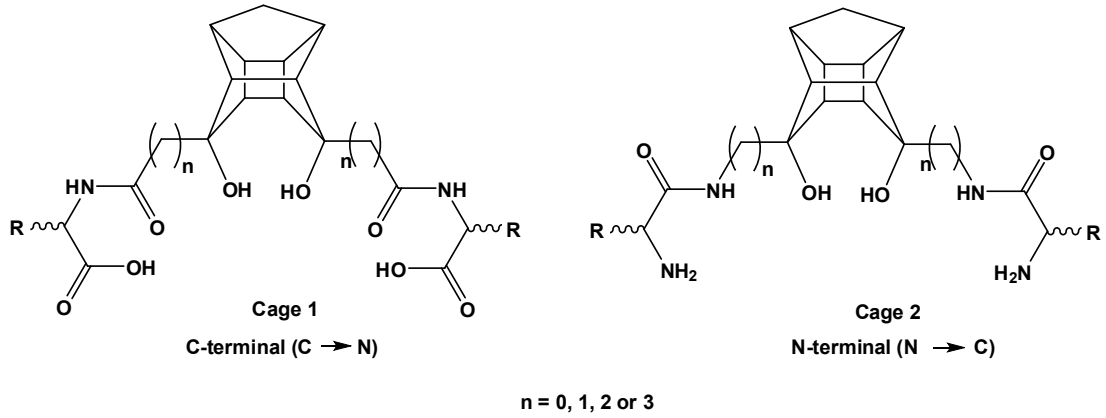


Figure 52. Model structure for the PCU-diol peptides to be investigated.

It was found that in all cases, n = 0, 1, 2 and 3, the N→C (R-cage**2**-R) compound, exhibited higher binding energies than C→N (R-cage**1**-R). Compounds with bulky hydrophobic amino acids, were found to be the most favoured when compared to less hydrophobic, polar, charged, basic and acidic residues in both carboxyl and amino terminal. For the C→N (R-cage**1**-R) connectivity, it was found that Glu and Asp residues gave weak binding energies while they showed improved binding in cage**2.** There was no conclusive trend in the chirality investigation however, more prominent energy differences were observed in some diastereomers with the carboxyl and amino terminal for n = 3 and n = 2, respectively.

Recommendations

Recently, the C-SA HIV PR crystal structure was solved thus, in future a similar study can be performed using this structure. In the current simulation studies, the inhibitors moved in the active site during the MD simulation which was due to the inhibitor not fitting tightly in the PR active site. It is really important to make sure that the inhibitor is tightly held in the active site during docking for better binding. Since, the comparisons of the flap movement for both unbound enzyme and the complexes were successful, a model can be built prior to synthesis. This model would help us to

understand how the new inhibitors behave in the active site thus, saving time and money required for synthesis.

Optimization of the parameters for the explicit solvation model is required since it was observed that the method used in this study required much longer simulation times. This approach should give results which are more closely related to the mode of action thus, the role of water molecules during catalysis can be investigated.

In future, prior to MD simulations the initial conformation of the inhibitors should resemble the conformations obtained in crystallographic studies. This would be useful in order to compare our PCU-based compounds with the approved antiviral drugs. To gain more insight into the structure-activity relationship for the PCU-based compounds, binding energies from the MD simulation should be calculated. The obtained binding energies will then be compared with the experimental IC_{50} and how they affect the protease movement. This approach will help us in understanding how the inhibitor behaves in the protease active site thus, helping in the design of more potent inhibitors which will either interact with the catalytic residues, with the flap residues or both.

Appendices

Appendix A: Chapter 3 supporting information

Appendix B: Chapter 3 input files for MD simulations

Appendix A

Chapter 3 supporting information

A.1 The chosen docked complex structures used in molecular dynamic simulations

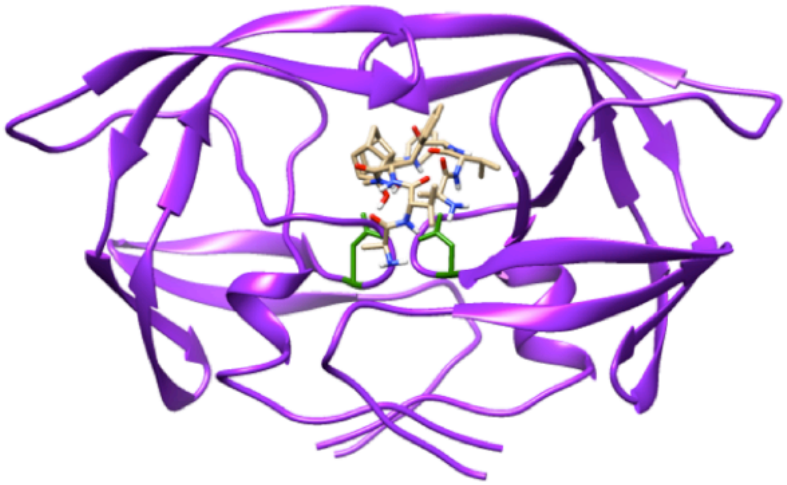


Figure 53. The chosen docked structure for PCU-diol peptide C-SA PR.

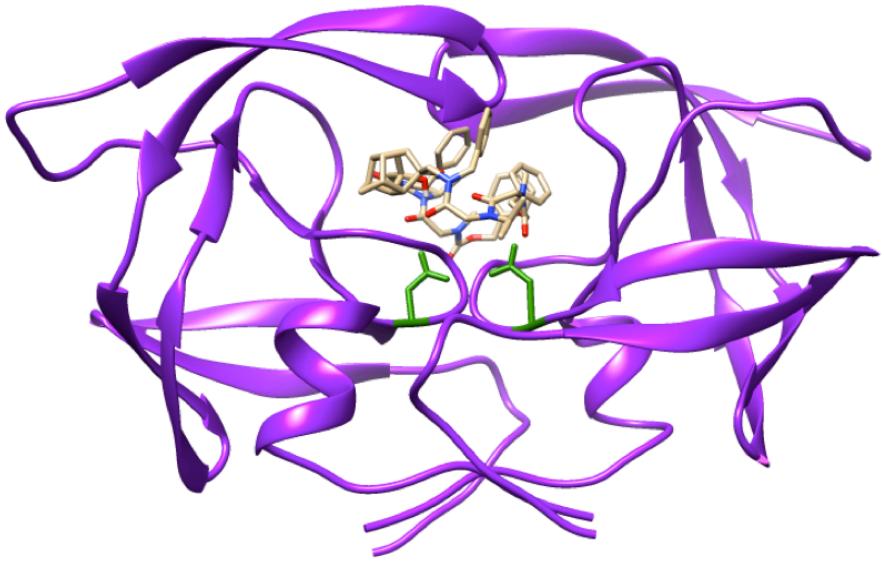


Figure 54. The chosen docked structure for PCU-diol peptoid C-SA PR.

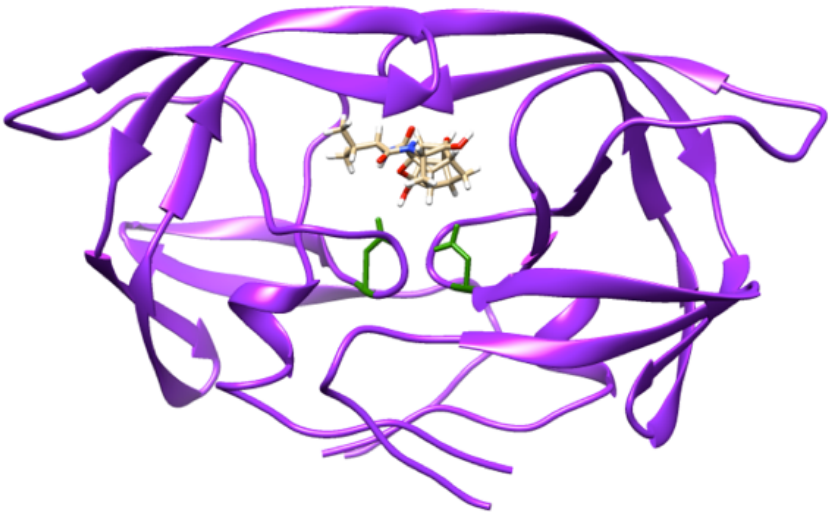


Figure 55. The chosen docked structure for PCU-ether peptide C-SA PR.

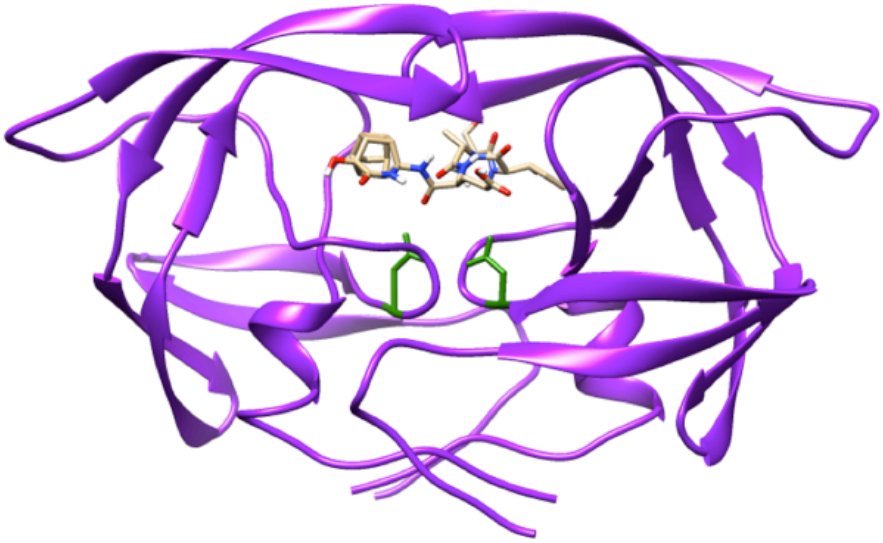


Figure 56. The chosen docked structure for PCU-lactam peptide C-SA PR.

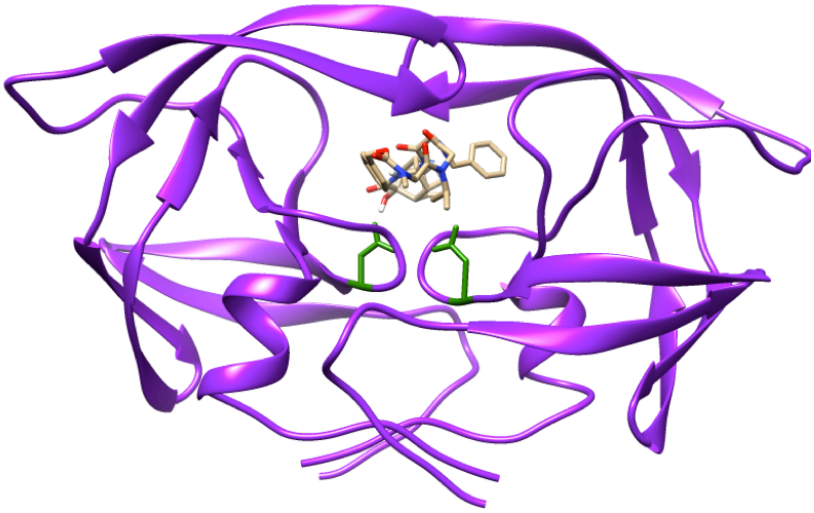


Figure 57. The chosen docked structure for PCU-lactam peptoid C-SA PR.

Appendix B

Chapter 3 input files for MD simulations

B.1 Input files for MD simulations in explicit solvation

The following in put files were used in this thesis.

Partial minimization input file

```
&cntrl
  imin   = 1,
  maxcyc = 2500,
  ncyc   = 750,
  ntb    = 1,
  ntr    = 1,
  cut    = 12.0,
 /
Hold the Protein fixed
500.0
RES 1 198
END
END
```

Full minimization input file

```
&cntrl
  imin   = 1,
  maxcyc = 200,
  ncyc   = 50,
  ntb    = 1,
  ntr    = 0,
  cut    = 12.0,
  Drms   = 0.0001,
/
```

Heating stage 1-6 input file

The temperature of the heating stage one to six were as follows (0-50)K, (50-100)K, (100-150)K, (150-200)K, (200-250)K, (250-300)K, respectively

```
&cntrl
  imin= 0,
  irest=0,
  NTX=1,
  ntb= 1,
  NTPR=500,
  NTWX=500,
  NTWR=500,
```

ntr=1,

Tempi=00.0,

Temp0=50.0,

NTT=3,

gamma_ln=1.0,

NTC=2,

NTF=2,

cut= 12.0,

nstlim=2500,

dt=0.002,

/

Keep Protein and inhibitor fixed with weak restraints

10.0

RES 1 198

END

END

Data input file

&cntrl

imin= 0,

irest=1,

NTX=7,

ntb=2,

ntp=1,

PRES0=1.0,

TAUP=2.0,

NTPR=500,

NTWX=500,

ntr=0,

Tempi=300.0,

Temp0=300.0,

NTT=3,

gamma_ln=1.0,

NTC=2,

NTF=2,

cut=12.0,

nstlim=250000,

dt=0.002

/

Input files for MD simulation in implicit solvation

The following in put files were used in this thesis.

Partial minimization input file

&cntrl

imin = 1,

maxcyc = 5000,

ncyc = 2500,

ntb = 0,

ntr = 1,

cut = 999.0,

IGB = 2

/

Hold the Protein fixed

500.0

RES 1 198

END

END

Full minimization input file

&cntrl

imin = 1,

maxcyc = 5000,

ncyc = 2500,

ntb = 0,

ntr = 0,

cut = 999.0,

IGB = 2,

```
 Drms   = 0.0001
/
```

Heating stages 1-6 input file

The temperature of the heating stage one to six were as follows (0-50)K, (50-100)K, (100-150)K, (150-200)K, (200-250)K, (250-300)K, respectively.

```
&cntrl
 imin= 0,
 irest=0,
 NTX=1,
 ntb= 0,
 igb = 2,
 NTPR=500,
 NTWX=500,
 NTWR=500,
 ntr=1,
 Tempi=00.0,
 Temp0=50.0,
 NTT=3,
 gamma_ln=1.0,
 NTC=2,
 NTF=2,
```

cut= 999.0,

nstlim=2500,

dt=0.002

/

Keep Protein and inhibitor fixed with weak restraints

10.0

RES 1 198

END

END

Data input file

&cntrl

imin= 0,

irest=1,

NTX=5,

ntb=0,

igb=2,

TAUP=2.0,

NTPR=500,

NTWX=500,

ntr=0,

Tempi=300.0,

```
 Temp0=300.0,
 NTT=3,
 gamma_ln=1.0,
 NTC=2,
 NTF=2,
 cut=999.0,
 nstlim=2500000,
 dt=0.002,
/
```

B.2 Data analysis parameters

The data analysis was done using ptraj found in Amber and the plots created by Origin software.

Measuring the distance between Cα Ile50 and Cα Ile149

trajin Data_10ns.mdcrd

distance DIST @2342 @778 out ptraj_distance.out

Measuring the distance between Cα Lys55 and CαLys154

trajin Data_10ns.mdcrd

distance DIST @2414 @850 out ptraj_K55.out

Measuring the distance between Cα Ile50 and Cβ Asp25

trajin Data_10ns.mdcrd

distance DIST @778 @407 out ptraj_monomerA.out

Measuring the distance between Cα Ile149 and Cβ Asp 124

trajin Data_10ns.mdcrd

distance DIST @2342 @1971 out ptraj_monomerB.out

Calculating the Flap A Cα RMSD

trajin Data_10ns.mdcrd

rms first out rmsd_CA_A01 :43-58@CA

Calculating the Flap B Cα RMSD

trajin Data_10ns.mdcrd

rms first out rmsd_CA_B01 :142-157@CA

Calculating RMSD for the backbone

trajin Data_10ns.mdcrd

rms first out rmsd @CA,C,N

Calculating the RMSF of the backbone

trajin Data_10ns.mdcrd

atomicfluct out RMSF :1-198 byres

Measuring the angle formed from Cα of G48-G49-I50 in monomer A

trajin Data_10ns.mdcrd

angle angle_nameA_01 @764 @771 @778 out angleA_01.dat time 1.0

Measuring the angle formed from Cα of G147-G148-G149 in Monomer B

trajin Data_10ns.mdcrd

angle angle_nameB_01 @2328 @2335 @2342 out angleB_01.dat time 1.0