UNIVERSITY OF KWAZULU-NATAL

THE MECHANISTIC MODELLING OF HIV-1 PROTEASE AND ITS NATURAL SUBSTRATES: A THEORETICAL PERSPECTIVE

2020

ZAINAB KEMI SANUSI

215079499

THE MECHANISTIC MODELLING OF HIV-1 PROTEASE AND ITS NATURAL SUBSTRATES: A THEORETICAL PERSPECTIVE

ZAINAB KEMI SANUSI

215079499

2020

A thesis submitted to the School of Pharmacy and Pharmacology, College of Health Science, University of KwaZulu-Natal, Westville, for the degree of Doctor of Philosophy in Health Science (Pharmaceutical chemistry).

This is the thesis in which the chapters are written as a set of discrete research publications, with an overall introduction and final summary. Most of the chapters has been published in internationally recognized, peer-reviewed journals.

This is to certify that the content of this thesis is the original research work of Zainab Kemi Sanusi.

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

Supervisor: Signed: ---------- Name: **Dr. G. E. M. Maguire** Date: 10/03/2020

Co-Supervisor: Signed: ---------- Name: **Prof. H. G. Kruger** Date: 10/03/2020

DEDICATION

This dissertation is proudly dedicated to my parents Alhaji. & Alhaja. Sanusi, for the unquantifiable values instilled in me. You two are the best.

ABSTRACT

An epidemic that has had profound impact on humanity both culturally and health-wise in recent decades is the Acquired immunodeficiency syndrome (AIDS) triggered by the Human immunodeficiency virus (HIV). The developments of drugs, impeding specific enzymes essential for the replication of the HIV-1 virus, has been a breakthrough in the treatment of the virus. These enzymes include the HIV-1 protease (PR), which is a significant degrading enzyme necessary for the proteolytic cleavage of the Gag and Gag-Pol polyproteins, needed for the maturation of viral protein. The catalytic mechanism of the HIV-1 PR of these polyproteins is a major subject of investigation over the past decades.

Most research on this topic explores the HIV-1 PR mechanism of action on its target as a stepwise general acid-base mechanism with little or no attention to the concerted process. Among the limitations of the stepwise reaction model is the presence of more than two transition state (TS) steps and this led to different views on the precise rate-determining step of the reaction as well as the protonation state of the catalytic aspartate in the active site of the HIV-1 PR. Likewise, consensus on the exact recognition mechanism of the natural substrates by HIV-1 PR is not forthcoming. The present study investigates the recognition approach and mechanism of reaction of the HIV-1 PR with its natural substrate by a means of computational models. It is intended to explain the cleavage mechanism of the reaction as a concerted process through the application of *in-silico* techniques. This is achieved by computing the activation energies and elucidating the quantum chemical properties of the enzyme-substrate system. An improved understanding of the mechanism will assist in the design of new HIV-1 PR inhibitors.

The molecular dynamics (MD) technique with hybrid quantum mechanics and molecular mechanics (QM/MM) method that includes the density functional theory (DFT) and Amber model were utilized to investigate the concerted hydrolysis process. Based on previous studies in our group involving concerted TS modeling, a six-membered ring TS pathway was first considered. This was achieved by employing a small model system and QM methods (Hartree-Fock and DFT) for the enzymatic mechanism of the HIV-1 PR. A general-acid base (GA/GB) model where both catalytic aspartate (Asp) groups are involved in the mechanism, and the water molecule at the active site attacks the natural substrate synchronously, was utilized. A new perspective arose from

the study where an acyclic concerted computational model offered activation energies closer to experiment observations than the six-membered ring model. Hence, the proposed concerted acyclic mechanism of the HIV-1 natural substrate within the entire protease was investigated using both multi-layered QM/MM "Our own N-layered Integrated molecular Orbital and molecular Mechanics" (ONIOM) theory and QM/MM MD umbrella sampling method.

A comprehensive review about experimental and theoretical results for the interactions between HIV PR and its natural substrates was presented. An important output in the present study is that the acyclic TS model barrier with one water molecule at the HIV-1 PR active site (DFT study), provides marginally, the most accurate activation energies. Similarly, the computational model demonstrated that optimum recognition specificity of the enzyme depends on structural details of the substrates as well as the number of amino acids in the substrate sequence (minimum P5-P5′ required). By modelling the entire enzyme—substrate system using a hybrid ONIOM QM/MM method, it was observed that although both subtype B and C-SA HIV-1 PR recognize and cleave at the scissile and non-scissile regions of the natural substrate sequence, the scissile region has a lower activation free energy. In all cases we found activation free energies that are in good agreement with experimental results. Also, the free energy profiles obtained from the umbrella sampling model were in absolute agreement with experimental *in vitro* HIV-1 PR hydrolysis data. The outcome of this investigations offers a plausible theoretical yardstick for the concerted enzymatic mechanism of the HIV-1 PRs that is pragmatic to related aspartate proteases and possibly other enzymatic processes.

Future studies on the reaction mechanism of HIV-1 PR and its natural substrate should encompass the use of advanced theoretical techniques aimed at exploring more than the energetics of the system. The prospect of integrated computational algorithms that does not involve cropped/partitioning/constraining or restraining model systems of the enzyme—substrate mechanism to accurately elucidate the HIV-1 PR catalytic process on natural substrates/inhibitors will be undertaken in our group. Computational investigations on the enzymatic mechanism of the HIV-1 PR—natural substrate involves fine-tuning the scissile amide bond strength through steric and electronic factors. This may lead to the development of potential substrate-based inhibitors with better potency and reduced toxicity.

ISIQEPHU

Ubhubhane olube nomthelela omkhulu ebuntwini bobabili ngokwemvelo nangokuqonda kwezempilo emashumini eminyaka amuva nje yi-Acquired immunodeficiency syndrome (AIDS) okubangelwa yi-Human immunodeficiency virus (HIV). Ukuthuthuka kwezidakamizwa, okufaka amandla ama-enzyme athile abalulekile ekuphindaphindweni kwegciwane le-HIV-1, kube yimpumelelo ekwelashweni kwaleli gciwane. La ma-enzyme afaka i-HIV-1 proteinase (PR), okuyi-enzyme ebalulekile eyonakalisayo edingekayo ekuhlanzeni kwe-protein ye-Gag ne-Gag-Pol, edingeka ekuvuthweni kweprotheni yegciwane. Indlela ebusayo ye-HIV-1 PR yalezi zi-polyprotein iyinto enkulu ephenywayo emashumini eminyaka edlule.

Ucwaningo oluningi ngalesi sihloko luhlola indlela esebenza ngayo ye-HIV-1 PR kulokho okukuhlosile njengenyathelo elisisekelo le-acid-base elisebenzayo ngaphandle kokunaka noma lengenayo inqubo ehlanganisiwe. Phakathi kokukhawulelwa kwemodeli yokusabela esezingeni eliphansi kukhona ubukhona bezinyathelo ezingaphezu kwezimbili zokuguqula isimo (TS) futhi lokhu kuholele ekubukweni okuhlukile esilinganisweni esinqunyiwe sokulinganisa sokuphendula kanye nesimo sokuhlasela sethonya elishukumisayo kulowo osebenzayo indawo ye-HIV-1 PR. Ngokunjalo, ukuvumelana mayelana nendlela ngqo yokuqashelwa kwezakhi zemvelo nge-HIV-1 PR akusondeli. Ucwaningo lwamanje luphenya indlela yokuqashelwa kanye nendlela yokusabela kwe-HIV-1 PR ngesakhiwo sayo esingokwemvelo ngezindlela zamamodeli wokuncintisana. Kuhloswe ukuchaza indlela ye-cleavage yokusabela njengenqubo ekhonjiwe ngokusebenzisa amasu we-in-silico. Lokhu kutholakala ngokuhlanganisa amandla we-activation amandla kanye nokucacisa izakhiwo zamakhemikhali we-quantum wohlelo lwangaphansi lwe-enzyme. Ukuqonda okungcono kwendlela ezokusiza ekwakhiweni kwama-inhibitors amasha we-HIV-1 PR.

Indlela esetshenziswayo yama-molecule (i-MD) ene-hybrid quantum mechanics kanye ne-molecule mechanics (QM / MM) efaka inqubo yokusizakala yokusebenza kwe-density theory (DFT) kanye ne-Amber model ukuphenya inqubo ekhonjiwe ye-hydrolysis. Ngokusekelwe kwizifundo zangaphambili eqenjini lethu ezibandakanya ukumodelwa kwe-TS ekhonjiwe, indlela eyindilinga eyisithupha yomgwaqo eyi-TS yaqala ukubhekwa. Lokhu kutholwe ngokusebenzisa uhlelo olusha lwemodeli nezindlela ze-QM (Hartree-Fock ne-DFT) ngomshini we-enzymatic we-

HIV-1 PR. Imodeli ejwayelekile ye-acid-(GA / GB) lapho amaqembu womabili we-catalytic aspartate (Asp) abandakanyeka khona emshinini, futhi i-molecule lamanzi esakhiweni esisebenzayo lihlasela i-substrate yemvelo ngokuvumelanisa, lalisetshenziswa. Kuqhamuke umbono omusha ocwaningweni lapho imodeli ye-acyclic ekhonjiwe yokuhlinzekwa kwamandla inika amandla okusebenzisa eduze nokuhlolwa okubonwayo kunasekuqaleni kwendandatho eyindandatho eyisithupha. Ngakho-ke, indlela ehlongozwayo ekhonjwe ngendlela ekhanyayo ye-HIV-1 substrate yemvelo kuyo yonke iprotease iphenyisisiwe kusetshenziswa ama-QM / MM amaningi ahlukaniswe ngama-Mechanics”(ONIOM) kanye ne-QM / MM MD isampula isambulela indlela.

Ukubuyekezwa okuphelele mayelana nemiphumela yokulinga kanye nemibhalo theory yokuxhumana phakathi kwe-HIV PR nezakhi zayo zemvelo kwalethwa. Umphumela obalulekile ocwaningweni lwamanje ukuthi isithintelo se-acyclic TS imodeli nge-mocule eyodwa yamanzi kwisiza esisebenzayo se-HIV-1 PR (i-DFT), sinikela ngamandla, amandla anembe kakhulu okusebenza. Ngokufanayo, imodeli yokuhlanganisa ibonise ukuthi ukuqashelwa okuphelele kwe-enzyme kuncike kwimininingwane yokwakheka kwama-substrates kanye nenani lama-amino acid ngokulandelana kwe-substrate (ubuncane be-P5-P5'). Ngokumodela yonke i-enzyme — uhlelo olusebenzisa uhlelo lwe-hybrid ONIOM QM / MM, kwaqapheleka ukuthi yize zombili izifunda ezingaphansi kwe-B ne-C-SA ye-HIV-1 PR zibona futhi zinamathele ezindaweni ezibucayi nezingasontekile zendlela yokulandelana engokwemvelo. isifunda esinomswakama sinamandla aphansi we-activation mahhala. Kuzo zonke izimo sithole amandla we-activation mahhala avumelane kahle nemiphumela yokuhlolwa. Futhi, amaphrofayili wamandla wamahhala atholakala kusampuli yesampuli ye-umbrella ayesesivumelwaneni ngokuphelele nedatha yokuhlolwa kwe-vitro HIV-1 PR hydrolysis. Umphumela walolu phenyo uhlinzeka ngokungenaphutha kwethiyori eyingqophamlando ye-enzymatic mechanism ye-HIV-1 PRs edlulele kumaphrotheni ahlobene ne-aspartate kanye nezinye izinqubo ze-enzymatic.

Izifundo zesikhathi esizayo mayelana nendlela yokusebenza kwe-HIV-1 PR kanye nengxenye yayo yemvelo kufanele ifake phakathi ukusetshenziswa kwamasu athuthukile we-theorytical okuhloswe ngawo ukuthola ngaphezu komfutho we-system. Ithemba lama-algorithms ahlanganisiwe wokubandakanya okungabandakanyanga okuhlanganisiwe / ukwahlukanisa /

ukuphoqelela noma ukuvimba izindlela eziyimodeli ze-enzyme-inqubo engaphansi yokwengeza ukucacisa ngokunembile inqubo yokulwa ne-HIV-1 PR kuzakhi zangaphansi zemvelo / ezinqandweni kuzokwenziwa eqenjini lethu. Uphenyo lwe-computational mayelana ne-enzymatic mechanism ye-HIV-1 PR-substrate yemvelo ifaka phakathi ukulungisa kahle amandla e-bond ayisihlanganisi nge-steric ne-elekthronikhi. Lokhu kungaholela ekwakhiweni kwama-inhibitors angaphansi komhlaba angaphansi nge-potency engcono nokunciphisa ubuthi.

DECLARATION

I, Zainab Kemi Sanusi, declare that:

1. The research reported in this thesis, except where otherwise indicated, 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 reference's sections.

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

Zainab Kemi Sanusi

Signed:

Date: 05/03/2020

LIST OF PUBLICATIONS

Publication 1:

From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets

Authors: Monsurat M. Lawal, **Zainab K. Sanusi,** Thavendran Govender, Glenn E. M. Maguire, Bahareh Honarparvar, and Hendrik G. Kruger.

Zainab K. Sanusi and Monsurat M. Lawal contributed equally to the design of the project and writing of the review.

Thavendran Govender, Glenn E. M. Maguire and Bahareh Honarparvar: **Co-supervisors**

Hendrik G. Kruger: **Supervisor**

Accepted in Current Medicinal Chemistry on 29th October 2018. DOI: https://doi.org/10.2174/0929867325666181113122900

Publication 2:

Unravelling the concerted catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease: a hybrid QM/MM study

Authors: Monsurat M. Lawal, **Zainab K. Sanusi,** Thavendran Govender, Gideon F. Tolufashe Glenn E. M. Maguire, Bahareh Honarparvar, and Hendrik G. Kruger.

Monsurat M. Lawal, is the main author of the research, **Zainab K. Sanusi** participated in the design, calculations and writing of the manuscript, while Gideon F. Tolufashe also contributed substantially in the area of calculations.

Thavendran Govender, Glenn E. M. Maguire and Bahareh Honarparvar: **Co-supervisors**

Hendrik G. Kruger: **Supervisor**

Accepted in Structural Chemistry on 23rd November 2018. Structural Chemistry volume 30, pages 409–417(2019)

Publication 3:

Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates

Authors: **Zainab K. Sanusi,** Monsurat M. Lawal, Thavendran Govender, Glenn E. M. Maguire, Bahareh Honarparvar, and Hendrik G. Kruger.

Zainab K. Sanusi, is the main author of the research, she participated in the design, calculations and writing of the manuscript, Monsurat M. Lawal participated in the design, calculations and writing of the manuscript.

Thavendran Govender, Bahareh Honarparvar and Hendrik G. Kruger: **Co-supervisors**

Glenn E. M. Maguire: **Supervisor**

Accepted in Journal of Physical Chemistry B on 8th July 2019. DOI: 10.1021/acs.jpcb.9b02207.

Publication 4:

Concerted hydrolysis mechanism of HIV-1 natural substrate against subtypes B and C-SA PR: Insight through Molecular Dynamics and Hybrid QM/MM studies

Authors: **Zainab K. Sanusi,** Monsurat M. Lawal, Thavendran Govender, Sooraj Baijnath, Tricia Naicker, Glenn E. M. Maguire, Bahareh Honarparvar, and Hendrik G. Kruger.

Zainab K. Sanusi, is the main author of the research, she participated in the design, calculations and writing of the manuscript, Monsurat M. Lawal also contributed substantially in the calculations and writing of the manuscript.

Thavendran Govender, Sooraj Baijnath, Tricia Naicker, Bahareh Honarparvar and Hendrik G. Kruger: **Co-supervisors**

Glenn E. M. Maguire: **Supervisor**

Accepted in Physical Chemistry Chemical Physics on 19th December 2019. Phys. Chem. Chem. Phys., 2020,22, 2530-2539.

Publication 5:

Exploring the Concerted Mechanistic Pathway for HIV-1 PR—Substrate Revealed by Umbrella Sampling Simulation

Authors: **Zainab K. Sanusi,** Monsurat M. Lawal, Pancham L. Gupta, Thavendran Govender, Sooraj Baijnath, Tricia Naicker, Glenn E. M. Maguire, Bahareh Honarparvar, Adrian E. Roitberg and Hendrik G. Kruger.

Zainab K. Sanusi, is the main author of the research, she participated in the design, calculations and writing of the manuscript, Monsurat M. Lawal and Pancham L. Gupta also contributed substantially in the calculations and writing of the manuscript.

Thavendran Govender, Sooraj Baijnath, Tricia Naicker, Bahareh Honarparvar and Hendrik G. Kruger: **Co-supervisors**

Glenn E. M. Maguire: **Supervisor**

Submitted to the Journal of Biomolecular Structure & Dynamics on 15th June 2020.

Publication 6:

The enzymatic mechanism of the Immunodeficiency Virus (HIV) protease: are we done yet? Human

Authors: Monsurat M. Lawal, **Zainab K. Sanusi,** Collins U. Ibeji, Thavendran Govender, Glenn E. M. Maguire, Bahareh Honarparvar, and Hendrik G. Kruger.

Monsurat M. Lawal is the main author of the research, she participated in the design, calculations and writing of the manuscript. **Zainab K. Sanusi** and Collins U. Ibeji contributed substantially in the area of calculations and writing part of manuscript.

Thavendran Govender, Glenn E. M. Maguire and Bahareh Honarparvar: **Co-supervisors**

Hendrik G. Kruger: **Supervisor**

This manuscript is highly confidential as we are busy with the synthetic part of its outcome.

Conference contributions

- Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates

  Authors: **Zainab K. Sanusi,** Monsurat M. Lawal, Thavendran Govender, Glenn E. M. Maguire, Bahareh Honarparvar and Hendrik G. Kruger

  Accepted for Poster presentation at the upcoming International Conference on Pure and Applied Chemistry *Mauritius, ES. July. 6 – 10, 2020*

- Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates

  Authors: **Zainab K. Sanusi,** Monsurat M. Lawal, Thavendran Govender, Glenn E. M. Maguire, Bahareh Honarparvar and Hendrik G. Kruger

  Poster presentation at the 11th CHPC National Meeting *Pretoria, SA. Dec. 4 – 8, 2018*

- Unravelling the concerted catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease

  Authors: Monsurat M. Lawal, **Zainab K. Sanusi,** Thavendran Govender, Gideon F. Tolufashe, Glenn E. M. Maguire, Bahareh Honarparvar and Hendrik G. Kruger

  Accepted for Poster presentation at the 10th International RAIS Conference on Social Sciences and Humanities *Princeton, New Jersey, USA. Aug. 22 – 23, 2018*

- Unravelling the concerted catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease

  Authors: Monsurat M. Lawal, **Zainab K. Sanusi,** Thavendran Govender, Gideon F. Tolufashe, Glenn E. M. Maguire, Bahareh Honarparvar and Hendrik G. Kruger

  Accepted for Oral presentation at International Conference on Recent Advances in Medical and Health Sciences (ICRAMHS) *New York, USA Aug. 16 – 17, 2018*

ACKNOWLEDGMENTS

My foremost appreciation, glorification and praise goes to my Creator, Almighty Allah, who has been my help in ages past and my hope in years to come. I say Alhamdhulillahi robil Al'ameen for seeing me through the thick and thin of this research work.

This thesis is the result of the dedicated effort of Prof. Gert Kruger, Dr. Glenn E. M. Maguire and Dr. Bahareh Honarparvar, who adroitly supervised the entire research work. I also appreciate Monsurat Lawal, Gideon Tolufashe, Collins Ibeji, and Thando Ntombela for their big, small and tiny assistance towards completing this research.

My overwhelming gratitude goes to my parents, Mr. L.K Sanusi and Mrs. I.Y Sanusi, my siblings (Hafeez Sanusi, Malik Sanusi, Minkahil Sanusi, Olaide Sanusi, Omolola Sanusi, Aminat Dauda) for their financial, moral and spiritual support. I love you so much and God bless you all.

My appreciation also goes to Prof. Thavi Govender, Dr. Tricia Naicker, other Catalysis and Peptide Research Unit (CPRU) principled investigators and the crew at large. My sincere thanks to: Adeola Shobo, Steve Ibijola, Ishaq Lawal, Samson Akpotu, Badru Abdulbaqi, Lekan Adisa for been a worthy brother, Monsurat Lawal for been a sister, best counselor, morale booster, companion, Ifeoma Anazonwu for her care, and the Nigerian Muslims around Westville for their friendliness, love, care and support.

My profound gratitude also goes to these distinct persons, my very own Sanusi families both near and far (I can only mention but a few), Badru Abdulbaqi, Aishat Akinyemi, Maryam Abdullahi, Muritala Ejalonibu, Fatima Ifeacho, Yakub Oluwagbemiga Seriki, Bode Ojugbele and family, for their love, motivation, prayers, support, well-wishes and care.

I also appreciate College of Health Science, University of KwaZulu-Natal and National Research Foundation, South Africa for this opportunity and financial support. I am also grateful to the Centre for High Performance Computing (www.chpc.ac.za) and UKZN Hippo cluster for computational resources.

LIST OF FIGURES

Chapter one;

Chapter two;

Chapter three;

Supporting information for Chapter seven;

LIST OF TABLES

LIST OF ACRONYMS

AIDS	Acquired Immune Deficiency Syndrome
$\Delta G^{\ddagger}$	Activation free energy
ANM	Anisotropic Network Model
AVB	Approximate valence bond
AMBER	Assisted Model Building with Energy Refinement
BSST	Biased sequence search threading
B3LYP	Becke3 Lee-Yang-Parr
CA	Capsid
WatC	Catalytic water
Cat D	Cathepsin D
COBRAMM Methods	Computations at Bologna Relating *Ab-initio* and Molecular Mechanics
CI	Configuration Interaction
DEE	Dead-end elimination
DFT	Density functional theory
EVB	Empirical valence bond
ES	Enzyme—substrate
FlAsH	Fluorescein arsenical hairpin
FDA	Food and Drug Administration
FEP	Free energy perturbation
HF	Hartree-Fock
HIV	Human Immunodeficiency Virus
HTLV-1	Human T-cell leukaemia virus
IN	Integrase
INT	Intermediate
IRC	Intrinsic reaction coordinates
KIEs	Kinetic Isotope Effects
LBHB	Low barrier hydrogen bond

MA	Matrix
MMFF	Merk Molecular Force Field
MD	Molecular dynamics
MM	Molecular mechanics
MM-PBSA	Molecular mechanics Poisson-Boltzmann surface area
MP	Moller-Plesset
MDR	Multidrug-resistant
NMR	Nuclear magnetic resonance
NC	Nucleocapsid
ONIOM	"Our own N-layered Integrated molecular Orbital and molecular Mechanics"
PME	Particle Mesh Ewald
Plm II	Plasmepsin II
PES	Potential energy surface
PR	Protease
PDB	Protein Data Bank
QM	Quantum mechanics
REMD	Replica-exchange Molecular Dynamics
RT	Reverse transcriptase
RH	RNAse H
RMS	Root Mean Square
SA	South Africa
p1	Spacer peptide 1
p2	Spacer peptide 2
S-groove	Substrate-groove
SI	Supporting Information
C-SA PR	Subtype C-South Africa Protease
TI	Thermodynamic integration
3-D	Three-dimensional
TD-DFT	Time-dependent density functional theory

TS	Transition state
TB	Tuberculosis
UFF	Universal Force Field
VDW	van der Waals
WT	Wild type

Table of Contents

CHAPTER ONE

INTRODUCTION

1.1 Background

A prolific epidemic that has had a profound health impact on humanity in this 21st century is the Acquired immunodeficiency syndrome (AIDS) caused by the Human immunodeficiency virus (HIV).[1] Since it was first recognized in the 80s,[2] HIV has grown from a single infection report to a global pandemic and remains one of the severest health challenges.[2] The HIV virus has two strains the HIV type 1 (HIV-1) and type 2 (HIV-2), where the former is the most predominant and virulent across the globe, the latter more common in West and Southern Africa.[1, 3] As a result of the number of people living with this infection both worldwide[4] and especially in South Africa (7.1 million individuals),[5] the quest for managing and treating this disease is therefore a high priority.

HIV attacks the host's immune system, which makes an individual vulnerable to opportunistic infection.[6] An intensive study of its structure and lifecycle led to the development of unrivaled anti-viral drugs inhibiting the three vital retroviral enzymes; protease (PR), reverse transcriptase (RT) and integrase (IN), necessary for replicating the virus.[7] The hydrolytic action of the protease on the natural amino acids substrate sequences results in processing the corresponding mature virion. The Gag proteolysis delivers substrates necessary for the processing of matrix (MA), nucleocapsid (NC), capsid (CA), spacer peptide 1 (p1), spacer peptide 2 (p2) and p6gag.[8] While Gag-Pol proteolysis delivers substrates that is required for RNAse H (RH), integrase (IN) and reverse transcriptase (RT).[9]

The HIV-1 protease (HIV-1 PR) is a degrading enzyme necessary for the proteolytic cleavage of the Gag and Gag-Pol polyproteins, which is required for the development of mature virion proteins.[10-12] One of the most active family of anti-retroviral drugs that is presently used in the treatment of HIV infection is the PI. These inhibitors are developed to mimic the Gag and Gag-Pol natural substrates of the HIV protease, which prevents the proteolytic activity of the enzyme and exert an inhibitory effect on replicating the HIV virus both *in vitro* and *in vivo*.[13, 14] A combination of antiretroviral therapy that includes PIs has resulted in a notable recovery of

immunodeficiency in patients and decreased AIDS related mortality.[15, 16] Nevertheless, failure to completely suppress the HIV replication leads to the development of viral resistance, this resistance to PIs does not only arise for the protease but also in some natural substrates of the protease.[13] Also, the long term use of the commercially available PIs causes adverse effects on the HIV patients.[17, 18] It is therefore necessary to develop new inhibitors with reduced side effects and toxicity, improved potency, long-lasting viral suppression and better oral bioavailability. Hence, the mechanism of action of the protease (proteolysis) on the Gag and Gag-Pol natural substrate is an important phenomenon.

Much research has focused on the mode of action of the HIV-1 PR on its natural substrates using various theoretical methods such as molecular dynamics (MD),[19-21] empirical valence bond (EVB),[22] a hybrid quantum mechanics/molecular mechanics (QM/MM),[23-26] MD QM/MM.[27, 28] Several experimental studies were also reported,[29-31] however, the exact mechanism of action still remains a subject of debate over to date.

Herein, the reaction mechanism and recognition of the Gag and Gag-Pol natural substrates sequence towards two HIV-1 PR subtypes (B and C-SA) was studied. This was carried out using computational tools in which an advance molecular simulation (Umbrella Sampling),[32] MD[19-21] and a multi-layer QM/MM ONIOM[33, 34] methods were harnessed to understand some related questions on the recognition of natural substrates and enzymatic catalysis of HIV-1 protease. The HIV-1 PR and its natural substrates which form the major focus of this research are discussed herein and theoretical tools/perspectives are introduced in the next chapter.

1.2 The HIV virus and lifecycle

The human immunodeficiency virus belongs to a family of retroviruses (lentivirus)[35] and can infect multiple cells in the human system such as dendritic cells and macrophages, however its major cell target is the CD4 lymphocyte also called the CD4+ or T cells.[36, 37] The HIV virus genes consist of specialized enzymes; protease, reverse transcriptase, integrase and two lone ribonucleic acid (RNA) strands that facilitate replication[38] (Figure 1).

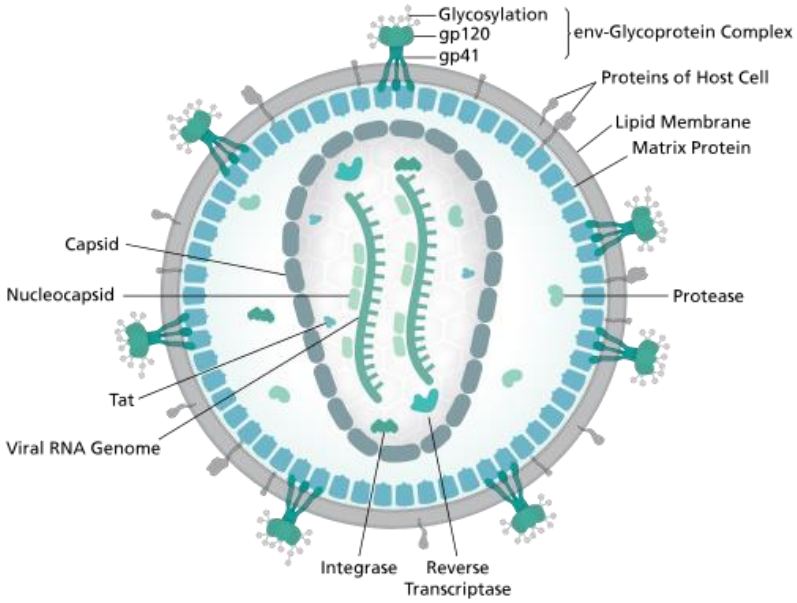


Figure 1. Diagrammatic representation of the HIV genome, (Reprinted with permission from ref 39. Copyright © 2014 S. Thomas).[39]

Once the HIV virus is manifested in the CD4+ cell, the infection can be classified into three steps: first the RNA strands are converted into deoxy-ribonucleic acid (DNA) by the viral reverse transcriptase.[40] By means of the integrase enzyme the newly formed viral double-stranded DNA genome is inserted into the host cell's DNA and this triggers the formation of new RNA, which is further transcribed into long protein chains that includes env, Gag and Gag-Pol precursors.[35, 41] Lastly, the protease cleaves the non-functional polyproteins into functional viral protein and together with the new immature RNA are assembled into viral particles, which then form the mature virus cells[42, 43] (Figure 2).

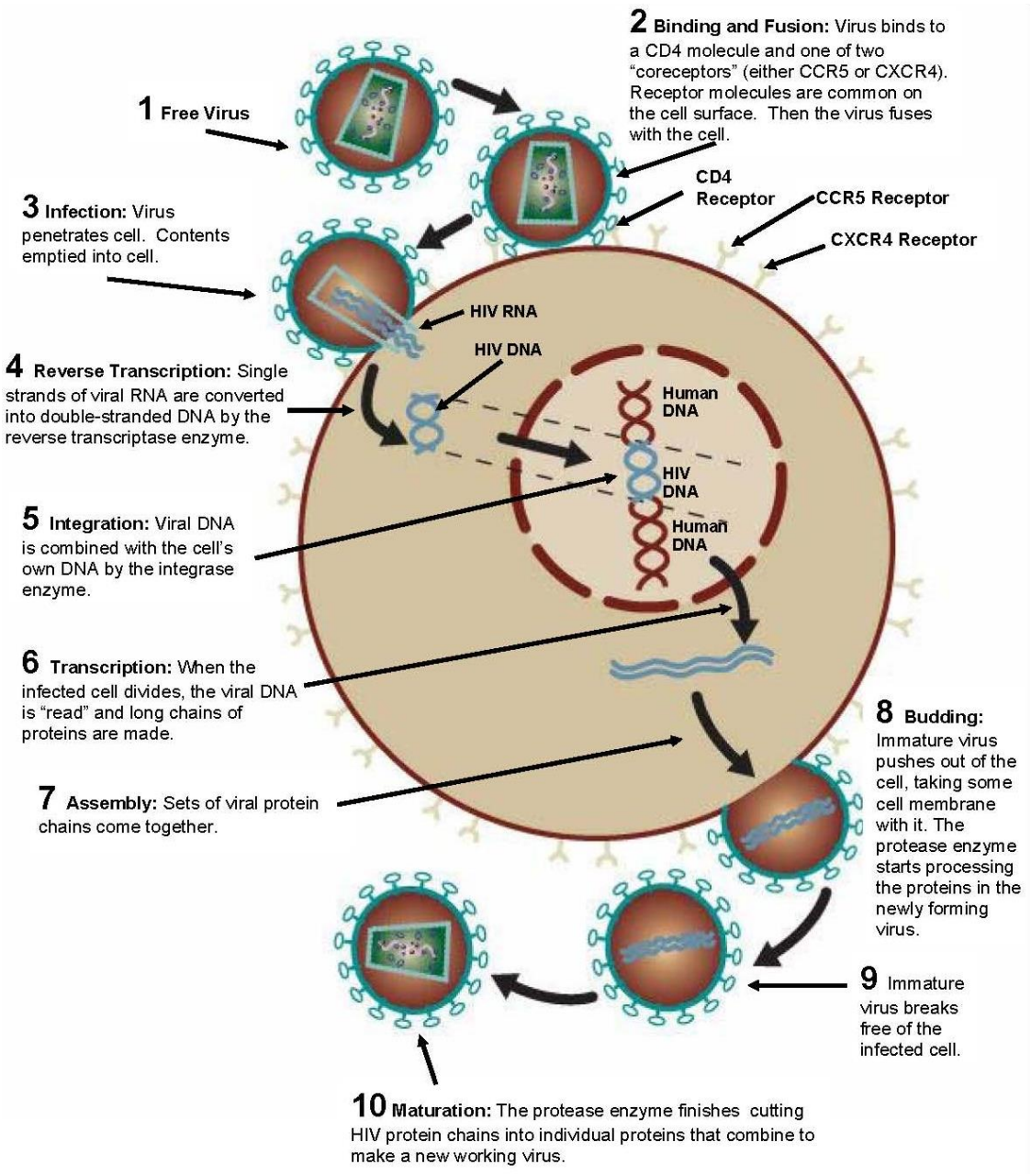


Figure 2. Schematic illustration of the HIV-1 life cycle. (Reprinted with permission from ref 44. Copyright © 2016 International Association of Providers of AIDS Care).[44]

1.3 The HIV-1 protease

HIV-1 PR belong to the family of aspartyl proteases[45] and as stated earlier, plays a vital role in the maturation of the HIV virus.[13, 46] Hence it serves as an important target for the development of anti-retroviral drugs.[10-12] HIV-1 protease is an identical C2-symmentric homodimer, which

normally contains 99-amino acids in each dimer in its active form[47] (Figure 3). However, variants of the subtype C-SA HIV-1 PR exist, for example, 100 (I36T↑T) and 101 (E35D↑G↑S) amino acid residues, that have been reported by our group.[48] The active site of the HIV-1 PR consist of a catalytic triad (Asp25-Thr26-Gly27) at each dimer, these are well conserved and similar in all HIV subtypes, the triads are enclosed by two β-sheet flaps.[49, 50] The catalytic aspartates interact directly with inhibitors and substrates and this has been associated with the enzymatic catalysis.[47] The PR enzyme has three regions/domain as shown in Figure 3 and are known as the dimerization or terminal domain, the flap and the core or active site domain.

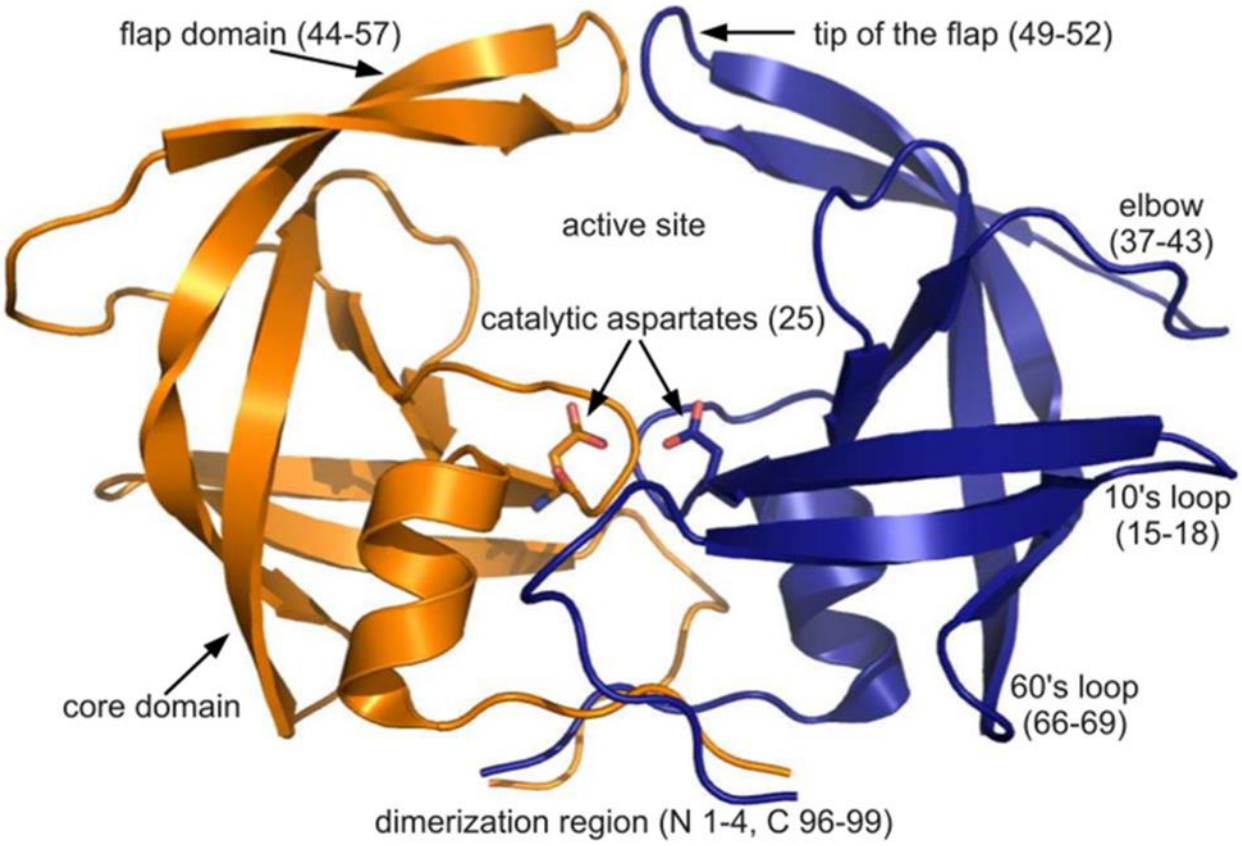


Figure 3. Structure of the HIV-1 PR showing its three domains.[51]

1.3.1 The HIV-1 PR regions

The dimerization or terminal domain is crucial in the dimer formation and stabilization of the active sites, this domain consists mainly of the termini of four stranded β-antiparallel sheets residues 1-4 and 95-99, a turn surrounding residue 4-9 and the helix residues 86-94 of each monomer.[52, 53]

The flap region comprises of two β-hairpin motifs with residues 43-58 of both monomers following the exposed loop 33-43 amino acid residues of each monomer. The HIV-1 PR flap exhibit structural rearrangement from closed to semi-open and wide-open form which encloses the active site and provides an important ligand/substrate binding interaction.[52, 54]

The active site cavity is also known as the core domain and is composed of four β-strand structures, this helps in stabilizing the dimer and the catalytic site.[16] The domain sequence comprises of compact 10-32 and 63-85 residues from each dimer. The conserved catalytic tripeptide Asp25-Thr26-Gly27 is positioned at the interface of the core domains of each monomer, the interface between the terminal and core regions is primarily composed of small hydrophobic residues.[52, 54]

1.4 HIV-1 natural substrate and substrate binding

The active site of the HIV-1 PR is composed of well-defined subsites where the substrates side chains can be accommodated while binding. The distinct subsites for each dimer are labelled S_1 to S_n and S_1' to S_n', and the corresponding amino acid side chains of the substrate cleaved during hydrolysis are labelled P1 to P_n and P_1' to P_n' starting from the scissile bond.[55] However, due to the C2 symmetric nature of the HIV-1 protease, the subsites S1 and S1' are similar and this is the same for all other subsites in the protease (Figure 4).[13, 47]

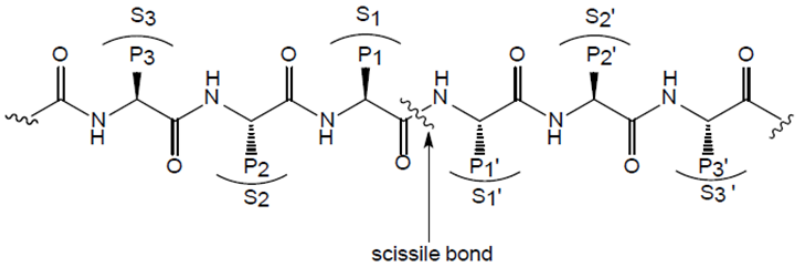


Figure 4. A representation of the PR substrate binding. P_1—P_n, P_1'—P_n' is a standard nomenclature used in representing amino acids of peptide residues, while protease binding sites are represented by S_1—S_n, S_1'—S_n' (Open access).[35, 47]

The hydrolytic cleavage of certain sections (scissile bond) in the Gag and Gag-Pol natural substrate precursors by HIV-1 PR leads to the release of specific functional polypeptide products, which are then processed into separated viral peptide units. HIV-1 PR recognizes and cleaves nine sites of these protein precursors which produces corresponding peptide segments of the natural substrates

(Table 1).[56-58] The mechanism of the natural substrate by the HIV-1 PR has been the premise of developing protease inhibitors for the treatment of HIV/AIDS for decades.[6, 59-61] This current research focuses on the interaction of the HIV-1 PR and its natural substrate and this will be extensively discussed in subsequent chapters.

Table 1. The non-homologous natural substrate gag and gag-pol polypeptides cleaved by the HIV-1 protease.[56, 57]

Peptide sequences cleavage domain	Natural substrate
Cleavage sites in Gag	
Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-Asn	MA-CA
Lys-Ala-Arg-Val-Leu*Ala-Glu-Ala-Met-Ser	CA-p2
Pro-Ala-Thr-Ile-Met*Met-Gln-Arg-Gly-Asn	p2-NC
Glu-Arg-Gln-Ala-Asn*Phe-Leu-Gly-Lys-Ile	NC-p1
Arg-Pro-Gly-Asn-Phe*Leu-Gln-Ser-Arg-Pro	p1-p6
Cleavage sites in Gag-Pol	
Val-Ser-Phe-Asn-Phe*Pro-Gln-Ile-Thr-Leu	TF-PR
Cys-Thr-Leu-Asn-Phe*Pro-Ile-Ser-Pro-Ile	PR-RT
Gly-Ala-Glu-Thr-Phe*Tyr-Val-Asp-Gly-Ala	RT-RH
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	RH-IN

The asterisk (*) denotes the scissile bond. Matrix-capsid; MA-CA, capsid-p2; CA-p2, p2-nucleopsid; p2-NC, nucleopsid-p1; NC-p1, *trans* frame peptide-protease; TF-PR, protease-reverse transcriptase; PR-RT, reverse transcriptase-RNAseH; RT-RH, RNAseH-integrase; RH-IN.[56, 57]

1.4.1 Scissile bond/cleavage domains

The cleavage sites within the Gag and Gag-Pol polypeptides that are recognised by the HIV-1 PR are provided in Figure 5. The ability of the HIV-1 protease to specifically recognize the natural substrates has been associated with the conserved shape of the peptide, substrate modulation, and the conformational adaptability of both substrate and protease that is interdependent.[24, 56, 57, 62, 63] A different opinion is that the HIV-1 PR recognises its natural substrate based on the conformational specificity of the PR for the Gag and Gag-Pol polyproteins sequence.[21]

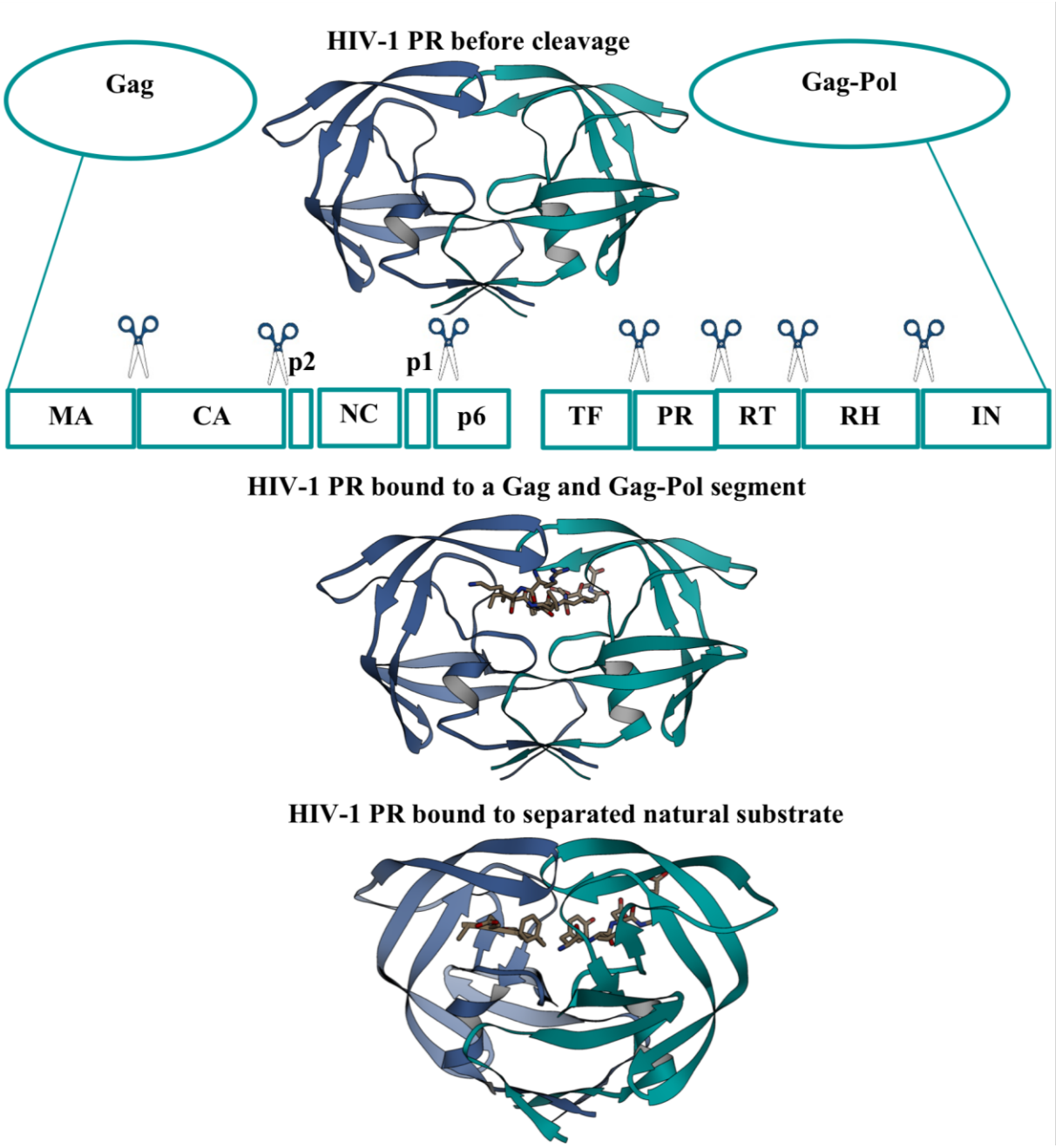


Figure 5. Recognition and reaction mechanism of the HIV-1 PR, each illustration is based on the PDB structures indicated: the apo HIV-1 PR; 1LV1,[64] substrate-bound HIV-1 PR; 1KJH[57] and broken scissile bond (the product peptides) bound to HIV-1 PR; 2NPH.[65] This structure was created using chimera.[66]

The non-homologous polypeptide sequences of each cleavage site have a distinguishing scissile bond (asterisk in Table 1) and this is the point of the HIV-1 PR proteolytic action. The design of the protease inhibitors (PIs) initiates from mimicking the scissile amino acids that occur within the cleavage domains especially the Phe*Pro scissile bonds (Table 1).[67]

The transition state (TS) modelling approach was utilized to better examine the natural substrate recognition pattern by HIV-1 PR (Chapter three and four). This study explores the activation barrier at different peptide bond units within the natural substrate sequences. This provided a comparative analysis of substrate specificity and recognition by the protease enzyme.

1.5 HIV-1 PR mutations

HIV-1 PRs have developed multiple mutations that show resistance to the commercially available anti-retroviral drugs over time.[68] The high rate of viral replication and lack of proofreading or self-correcting of the reverse transcriptase contributes considerably to a build-up of drug resistance of this protease.[68] Genome errors occur as the reverse transcriptase create the DNA virus from the RNA, coupled with the replication rate, new drug resistance generations of the virus can begin to outnumber the non-resistance virus.[68, 69] Amino acids insertion, deletion and substitution that occur in the virus sequence is as a result of mutations that arise through resistance.[70] The new mutant will also survive better in the presence of the same inhibitor/drug. The structure of the HIV PRs are influenced by these mutations, which leads to conformational and steric hindrance of the existing PIs, hence the need to continuously design and synthesis new inhibitors.[71-73]

HIV-1 subtype B differs structurally from the C-SA PR due to the eight-point mutations that occur away from the active site of the wild-type subtype B.[70] In addition, other mutations were discovered to occur from the C-SA PR amongst them are I36T↑T (with 100 amino acids in each monomer) and E35D↑G↑S PRs (with 101 amino acids in each monomer).[48]

1.6 HIV-1 PR catalytic mechanism

In designing the HIV-1 PIs the first approach used, and still used today, is mimicking the natural substrate, but with a structural changes that inhibit or stop the protease function.[35, 74] Hence, the HIV-1 protease interaction mechanism with its natural substrates is one of the most researched enzymes—substrate representatives for hydrolytic systems.[75] Experimental studies have been able to show some factors like the protonation state of the catalytic Asp, the role of structural water and identification of the catalytic triad (Asp25-Thr26-Gly27).[23, 41] Despite all that is known about the importance of the PR family and HIV-1 PR, its catalytic mechanism remains contentious, there is

no consensus on which is the correct/exact mechanism.[49, 58] The HIV-1 PR hydrolysis is usually investigated as a stepwise general-acid general-base mechanism (GA/GB) and this involves a water molecule and both protonated and unprotonated catalytic Asp at the active site of the enzyme.[27] Another proposed mechanism is the nucleophilic model whereby water molecule is absent in the calculations,[25] this approach appears to be fading as researchers continue to unveil the important role of water in HIV-1 protease catalysis as discovered through X-ray crystallographic studies.[64, 65, 76]

For the present study, both subtype B and C-SA HIV-1 PR enzymatic mechanistic pathways were investigated as a concerted general acid-base process (Chapters three to six). Whereby, the nucleophilic attack of water as well as the general acid-base contributions of the catalytic aspartates occurs simultaneously. This mechanism was initially proposed following an experimental study in 1991 by Jakólski[31] *et al.*, few computational studies[19, 21, 23, 25] have also attempted to unravel the mechanism with little progress. Since several enzymatic and chemical reactions that make use of a concerted mechanistic model were studied by our group,[77-83] a concerted transition state model was also applied to describe the subtype B/C-SA PR—natural substrate catalytic interactions using different computational techniques. Detailed results are presented in chapters three to six, the knowledge obtained during this research enable us to design new substrate-based inhibitors.

Aim and objectives

The overall scope of this study was to investigate the recognition mechanism of all the natural substrates by HIV-1 PR and subsequent catalytic interactions between them with the aid of computational applications. This is accomplished by the following objectives and are streamed down in ensuing chapters.

- **Chapter One**: Introductory background of the study.
- **Chapter two**: A computational background study.
- **Chapter three**: Literature review on recognition and catalysis of natural substrates by HIV-1 PR.

- **Chapter four**: To elucidate the proposed concerted six-membered ring transition state mechanism of the natural substrate within the entire enzyme using hybrid ONIOM QM/MM method. To analyze variations in relation to the earlier result (small model).
- **Chapter five**: To compare the activation energies at different peptide bonds other than the scissile region within two natural substrate sequences each from the Gag and Gag-Pol polypeptide. To investigate if the natural substrate recognition pattern is the same for both subtype B and C-SA HIV-1 PR using the hybrid QM/MM ONIOM method.
- **Chapter six**: To investigate the activation energies of all the natural substrates sequences for the same HIV-1 PR (subtype B and C-SA) using combine MD and ONIOM approach.
- **Chapter seven**: To elucidate the proposed concerted six-membered ring transition state mechanism of the natural substrates by subtype C-SA HIV-1 PR using the Umbrella Sampling method.
- **Chapter eight**: Overall conclusion of the research outcome.

Thesis outline

The dissertation is presented in a manuscript format and each chapter is directed in addressing one or more research questions. The first and last chapters contain a general introduction and an overall conclusion for the whole study. An outline is therefore highlighted below.

Chapter one gives an introductory overview of the dissertation and the main direction of the study is emphasised.

Chapter two introduces the different computational techniques accessible and the computational method used is highlighted.

Chapter three presents the literature review on recognition pattern and hydrolysis of natural substrates by HIV-1 PR.

Chapter four includes the application of ONIOM to investigate substrate hydrolysis by HIV-1 PR using an exact natural substrate.

Chapter five presents substrate recognition by both subtype B and C-SA HIV-1 PR using ONIOM approach, investigation of the recognition pattern and specificity of the PRs for cleavage at a specific amide bond was extensively explored.

Chapter six presents the concerted hydrolysis of the HIV-1 natural substrates against subtypes B and C-SA through combine MD simulations and ONIOM technique.

Chapter seven explores the concerted acyclic mechanism of the HIV-1 PR proposed through Umbrella sampling technique.

Chapter eight presents the overall conclusion on the research outcome.

References

1. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.-C., and Montagnier, L. (1983) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV, *Nature 312*, 767-768.
2. Gallo, R. C., Sarin, P. S., Gelmann, E., Robert-Guroff, M., Richardson, E., Kalyanaraman, V., Mann, D., Sidhu, G. D., Stahl, R. E., and Zolla-Pazner, S. (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS), *Science 220*, 865-867.
3. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987) Genome organization and transactivation of the human immunodeficiency virus type 2, *Nature 326*, 662-669.
4. UNIADS. (2018) AIDS by the number, http://www.unaids.org/en/resources/fact-sheet.
5. UNIADS. (2018), http://www.unaids.org/en/regionscountries/countries/southafrica.
6. Potempa, M., Lee, S.-K., Wolfenden, R., and Swanstrom, R. (2015) The triple threat of HIV-1 protease inhibitors, In *The Future of HIV-1 Therapeutics*, p. 203-241, Springer.
7. Bon, D., Stephan, C., Keppler, O., and Herrmann, E. (2012) Viral dynamic model of antiretroviral therapy including the integrase inhibitor raltegravir in patients with HIV-1, *BIOMATH 1* (1), p.1209251.
8. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., and Rosenberg, M. (1987) Human immunodeficiency virus protease expressed in escherichia coli exhibits autoprocessing and specific maturation of the gag precursor, *Proceedings of the National Academy of Sciences 84*, 8903-8906.
9. Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C.-T., Lumma, P. K., Freidinger, R. M., Veber, D. F., and Sigal, I. S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins, *Biochemical and Biophysical Research Communications 156*, 297-303.
10. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R., Scolnick, E. M., and Sigal, I. S. (1988) Active human immunodeficiency virus protease is required for viral infectivity, *Proceedings of the National Academy of Sciences 85*, 4686-4690.
11. Crawford, S., and Goff, S. P. (1985) A deletion mutation in the 5'part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins, *Journal of Virology 53*, 899-907.
12. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression, *Nature 331*, 280.
13. Clavel, F., and Mammano, F. (2010) Role of Gag in HIV resistance to protease inhibitors, *Viruses 2*, 1411-1426.
14. Liu, H., Müller-Plathe, F., and van Gunsteren, W. F. (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease, *Journal of Molecular Biology 261*, 454-469.
15. Pomerantz, R. J., and Horn, D. L. (2003) Twenty years of therapy for HIV-1 infection, *Nature Medicine 9*, 867-873.
16. Castro, H. C., Abreu, P. A., Geraldo, R. B., Martins, R. C., dos Santos, R., Loureiro, N. I., Cabral, L. M., and Rodrigues, C. R. (2011) Looking at the proteases from a simple perspective, *Journal of Molecular Recognition 24*, 165-181.

17. Vacca, J. P., and Condra, J. H. (1997) Clinically effective HIV-1 protease inhibitors, *Drug Discovery Today 2*, 261-272.
18. Martinez-Cajas, J. L., and Wainberg, M. A. (2007) Protease inhibitor resistance in HIV-infected patients: molecular and clinical perspectives, *Antiviral Research 76*, 203-221.
19. Chatfield, D. C., and Brooks, B. R. (1995) HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation, *Journal of the American Chemical Society 117*, 5561-5572.
20. Trylska, J., Bała, P., Geller, M., and Grochowski, P. (2002) Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease, *Biophysical Journal 83*, 794-807.
21. Perez, M., Fernandes, P., and Ramos, M. (2010) Substrate recognition in HIV-1 protease: a computational study, *The Journal of Physical Chemistry B 114*, 2525-2532.
22. Bjelic, S., and Åqvist, J. (2006) Catalysis and linear free energy relationships in aspartic proteases, *Biochemistry 45*, 7709-7723.
23. Chatfield, D. C., Eurenius, K. P., and Brooks, B. R. (1998) HIV-1 protease cleavage mechanism: a theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential, *Journal of Molecular Structure: THEOCHEM 423*, 79-92.
24. Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006) Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate, *Journal of Virology 80*, 3607-3616.
25. Park, H., Suh, J., and Lee, S. (2000) Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism, *Journal of the American Chemical Society 122*, 3901-3908.
26. Kipp, D. R., Hirschi, J. S., Wakata, A., Goldstein, H., and Schramm, V. L. (2012) Transition states of native and drug-resistant HIV-1 protease are the same, *Proceedings of the National Academy of Sciences 109*, 6543-6548.
27. Trylska, J., Grochowski, P., and McCammon, J. A. (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease, *Protein Science 13*, 513-528.
28. Kožíšek, M., Bray, J., Řezáčová, P., Šašková, K., Brynda, J., Pokorná, J., Mammano, F., Rulíšek, L., and Konvalinka, J. (2007) Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants, *Journal of Molecular Biology 374*, 1005-1016.
29. Northrop, D. B. (2001) Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases, *Accounts of Chemical Research 34*, 790-797.
30. Hyland, L. J., Tomaszek Jr, T. A., and Meek, T. D. (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism, *Biochemistry 30*, 8454-8463.
31. Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B., and Wlodawer, A. (1991) Structure at 2.5-. ANG. resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor, *Biochemistry 30*, 1600-1609.
32. Kästner, J. (2011) Umbrella sampling, *Wiley Interdisciplinary Reviews: Computational Molecular Science 1*, 932-942.

33. Dapprich, S., Komáromi, I., Byun, K. S., Morokuma, K., and Frisch, M. J. (1999) A new ONIOM implementation in Gaussian98. part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives, *Journal of Molecular Structure: THEOCHEM 461*, 1-21.
34. Morokuma, K. (2002) New challenges in quantum chemistry: quests for accurate calculations for large molecular systems, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 360*, 1149-1164.
35. Brik, A., and Wong, C.-H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & Biomolecular Chemistry 1*, 5-14.
36. Pomerantz, R. J. (2002) HIV: a tough viral nut to crack, *Nature 418*, 594-595.
37. Turner, B. G., and Summers, M. F. (1999) Structural biology of HIV, *Journal of Molecular Biology 285*, 1-32.
38. Kurth, R., and Bannert, N. (2010) Retroviruses: molecular biology, genomics and pathogenesis, *Horizon Scientific Press, 9* (7), p.867.
39. Thomas, S. (2014) Diagram of HIV virion.
40. Shehu-Xhilaga, M., and Oelrichs, R. (2009) Basic HIV virology, *HIV Management in Australasia*, 9-18.
41. Mildner, A. M., Rothrock, D. J., Leone, J. W., Bannow, C. A., Lull, J. M., Reardon, I. M., Sarcich, J. L., Howe, W. J., and Tomich, C.-S. C. (1994) The HIV-1 protease as enzyme and substrate: mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic properties, *Biochemistry 33*, 9405-9413.
42. Mitsuya, H., Yarchoan, R., and Broder, S. (1990) Molecular targets for AIDS therapy, *Science 249*, 1533-1544.
43. Katz, R. A., and Skalka, A. M. (1994) The retroviral enzymes, *Annual Review of Biochemistry 63*, 133-173.
44. InfoNet, T. A. (2013) HIV Life Cycle [ONLINE].
45. Promsri, S., Chuichay, P., Sanghiran, V., Parasuk, V., and Hannongbua, S. (2005) Molecular and electronic properties of HIV-1 protease inhibitor C 60 derivatives as studied by the ONIOM method, *Journal of Molecular Structure: THEOCHEM 715*, 47-53.
46. Ashorn, P., McQuade, T. J., Thaisrivongs, S., Tomasselli, A. G., Tarpley, W. G., and Moss, B. (1990) An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection, *Proceedings of the National Academy of Sciences 87*, 7472-7476.
47. Reetz, M. T., Merk, C., and Mehler, G. (1998) Preparation of novel HIV-protease inhibitors, *Chemical Communications, 19*, 2075-2076.
48. Maseko, S. B., Natarajan, S., Sharma, V., Bhattacharyya, N., Govender, T., Sayed, Y., Maguire, G. E. M., Lin, J., and Kruger, H. G. (2016) Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies, *Protein Expression and Purification 122*, 90-96.
49. Silva, A. M., Cachau, R. E., Sham, H. L., and Erickson, J. W. (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease, *Journal of Molecular Biology 255*, 321-340.
50. Robbins, A. H., Coman, R. M., Bracho-Sanchez, E., Fernandez, M. A., Gilliland, C. T., Li, M., Agbandje-McKenna, M., Wlodawer, A., Dunn, B. M., and McKenna, R. (2010) 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 66*, 233-242.
51. Venkatakrishnan, B., Palii, M.-L., Agbandje-McKenna, M., and McKenna, R. (2012) Mining the protein data bank to differentiate error from structural variation in clustered static structures: an examination of HIV protease, *Viruses 4*, 348-362.
52. Painter, G. R., Almond, M. R., Mao, S., and Liotta, D. C. (2004) Biochemical and mechanistic basis for the activity of nucleoside analogue inhibitors of HIV reverse transcriptase, *Current Topics in Medicinal Chemistry 4*, 1035-1044.
53. Briz, V., Poveda, E., and Soriano, V. (2006) HIV entry inhibitors: mechanisms of action and resistance pathways, *Journal of Antimicrobial Chemotherapy 57*, 619-627.
54. Dau, B., and Holodniy, M. (2009) Novel targets for antiretroviral therapy, *Drugs 69*, 31-50.
55. Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998) Molecular and biotechnological aspects of microbial proteases, *Microbiology and Molecular Biology Reviews 62*, 597-635.
56. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000) How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of HIV-1 protease1, *Journal of Molecular Biology 301*, 1207-1220.
57. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, *Structure 10*, 369-381.
58. Sussman, F., Villaverde, M. C., L Dominguez, J., and Danielson, U. H. (2013) On the active site protonation state in aspartic proteases: implications for drug design, *Current Pharmaceutical Design 19*, 4257-4275.
59. Pawar, S. A., Jabgunde, A. M., Govender, P., Maguire, G. E. M., Kruger, H. G., Parboosing, R., Soliman, M. E., Sayed, Y., Dhavale, D. D., and Govender, T. (2012) Synthesis and molecular modelling studies of novel carbapeptide analogs for inhibition of HIV-1 protease, *European Journal of Medicinal Chemistry 53*, 13-21.
60. Makatini, M. M., Petzold, K., Arvidsson, P. I., Honarparvar, B., Govender, T., Maguire, G. E. M., Parboosing, R., Sayed, Y., Soliman, M. E., and Kruger, H. G. (2012) Synthesis, screening and computational investigation of pentacycloundecane-peptoids as potent CSA-HIV PR inhibitors, *European Journal of Medicinal Chemistry 57*, 459-467.
61. Debouck, C. (1992) The HIV-1 protease as a therapeutic target for AIDS, *AIDS Research and Human Retroviruses 8*, 153-164.
62. Özen, A., Haliloğlu, T., and Schiffer, C. A. (2011) Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope, *Journal of Molecular Biology 410*, 726-744.
63. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease, *Chemistry & Biology 11*, 1333-1338.
64. Kumar, M., Kannan, K., Hosur, M., Bhavesh, N. S., Chatterjee, A., Mittal, R., and Hosur, R. (2002) Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations, *Biochemical and Biophysical Research Communications 294*, 395-401.
65. Das, A., Mahale, S., Prashar, V., Bihani, S., Ferrer, J.-L., and Hosur, M. (2010) X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short

ionic hydrogen bond SIHB with catalytic aspartate, *Journal of the American Chemical Society 132*, 6366-6373.
66. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.
67. Abdel-Rahman, H. M., Al-karamany, G. S., El-Koussi, N. A., Youssef, A. F., and Kiso, Y. (2002) HIV protease inhibitors: peptidomimetic drugs and future perspectives, *Current Medicinal Chemistry 9*, 1905-1922.
68. Freeman, S., and Herron, J. C. (2013) Evolutionary Analysis: pearson new international edition, *Pearson Higher Education*, *9* (7), p.867.
69. Race, E. (2001) Cross-resistance within the protease inhibitor class, *Antiviral Therapy 6*, 29-36.
70. Mosebi, S., Morris, L., Dirr, H. W., and Sayed, Y. (2008) Active-site mutations in the South African human immunodeficiency virus type 1 subtype C protease have a significant impact on clinical inhibitor binding: kinetic and thermodynamic study, *Journal of Virology 82*, 11476-11479.
71. Wensing, A. M., van Maarseveen, N. M., and Nijhuis, M. (2010) Fifteen years of HIV protease inhibitors: raising the barrier to resistance, *Antiviral Research 85*, 59-74.
72. Weber, I. T., and Agniswamy, J. (2009) HIV-1 protease: structural perspectives on drug resistance, *Viruses 1*, 1110-1136.
73. Agniswamy, J., Shen, C.-H., Aniana, A., Sayer, J. M., Louis, J. M., and Weber, I. T. (2012) HIV-1 protease with 20 mutations exhibits extreme resistance to clinical inhibitors through coordinated structural rearrangements, *Biochemistry 51*, 2819-2828.
74. Perez, M., Fernandes, P., and Ramos, M. (2007) Drug design: new inhibitors for HIV-1 protease based on nelfinavir as lead, *Journal of Molecular Graphics and Modelling 26*, 634-642.
75. Krzemińska, A., Moliner, V., and Świderek, K. (2016) Dynamic and electrostatic effects on the reaction catalyzed by HIV-1 protease, *Journal of the American Chemical Society 138*, 16283-16298.
76. Prashar, V., Bihani, S., Das, A., Ferrer, J.-L., and Hosur, M. (2009) Catalytic water co-existing with a product peptide in the active site of HIV-1 protease revealed by X-ray structure analysis, *PloS One 4* (11), p.e7860.
77. Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E. M., Kruger, H. G., and Honarparvar, B. (2017) Computational model for the acylation step of the β-lactam ring: potential application for l, d-transpeptidase 2 in mycobacterium tuberculosis, *Journal of Molecular Structure 1128*, 94-102.
78. Kruger, H. G., Mdluli, P., Power, T. D., Raasch, T., and Singh, A. (2006) Experimental and computational studies of the regioselective protection of hydantoins using anhydride, *Journal of Molecular Structure: THEOCHEM 771*, 165-170.
79. Singh, T., Kruger, H. G., Bisetty, K., and Power, T. D. (2012) Theoretical study on the formation of a pentacyclo-undecane cage lactam, *Computational and Theoretical Chemistry 986*, 63-70.
80. Lawal, M. M., Govender, T., Maguire, G. E. M., Honarparvar, B., and Kruger, H. G. (2016) Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study, *Journal of Molecular Modeling 22*, 235-246.

81. Lawal, M. M., Govender, T., Maguire, G. E. M., Kruger, H. G., and Honarparvar, B. (2018) DFT study of the acid-catalyzed esterification reaction mechanism of methanol with carboxylic acid and its halide derivatives, *International Journal of Quantum Chemistry 118*, 25497-25508.
82. Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E., and Kruger, H. G. (2013) Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity, *Chemical Reviews 114*, 493-537.
83. Rauf, S. M. A., Arvidsson, P. I., Albericio, F., Govender, T., Maguire, G. E. M., Kruger, H. G., and Honarparvar, B. (2015) The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study, *Organic & Biomolecular Chemistry 13*, 9993-10006.

CHAPTER TWO

Computational Background

1.1 Computer-aided drug design

Structure-based drug design has been a leading tool in HIV-1 protease inhibitors and has generated many positive results.[1-3] Drug design with the aid of computer processes are proving valuable in this area, from modelling the interactions of potential drugs and drug targets to analysing experimental results.[4, 5] Due to improved computer hard and soft-ware, computer-aided methods are supposed to be more cost effective and less time consuming than experimental methods, but caution against unrealistic expectations should be maintained.[4] For highly mutative enzymes such as HIV, time is a significant factor and the computational approach has proven to be a useful complimentary tool to experimental methods.[6-8] This technique utilizes computational chemistry in discovering new potential drugs and their targets and to give molecular insight into inhibitor-enzyme interactions that are not readily accessible from experimental methods and. [6-8]

1.2 Computational chemistry

Computational chemistry can be defined as the mathematical description of chemistry which covers a wide range of theoretical fields including; quantum mechanics, minimization of energy, molecular mechanics and dynamics and conformational analysis.[9, 10] Primarily these approaches with other computer-based models are used to calculate the molecular interactions and behaviour of chemical as well as enzymatic systems. These techniques have been coupled with chromatographic, spectroscopic and other experimental approaches to study drug action and inhibition such as; antimalaria,[11-13] antiretroviral,[4, 14] antituberculotic,[15] anticancer[16-19] and antibiotic drugs.[20-22] Studies on the inhibition of pathogens predominantly antiretroviral drugs have been skilfully addressed in our research group.[4, 23-27]

Computational approaches can be utilized to identify chemical processes using parameterized advanced software to solve challenges *via* quantum mechanics (QM), molecular mechanics (MM), hybrid QM/MM and or Molecular dynamic (MD) methods.[28] Reaction mechanisms yet to be

established by experimental kinetic studies can be tested through a series of electronic structure calculations. The combination of spectroscopic methods, molecular modelling and enzymatic drug design was extensively reviewed on recent advances of computational and theoretical chemistry and the rest of this chapter describes an overview of the computational methods applied in this study.[4, 29]

1.3 Theoretical models

Theoretical models are used to describe the chemical behaviour of molecular systems through a set of approximations, these algorithms are applied to atomic orbitals to compute energies, frequencies and perform geometry optimization of molecules.[28-30] Computational methods employed in chemistry are basically classified as quantum and non-quantum mechanical methods.[30]

1.3.1 Quantum mechanics

The QM mechanics also known as electronic structure theories is aimed at solving the 1926 Erwin Schröndinger equation usually expressed in terms of a wave function[31, 32] to understand the properties molecules (see equation 1).[30] This theory is costly considering the computational resources and time but more accurate, hence, practical for smaller systems.[4, 33, 34] There are several methods of QM theory available to researchers and these are summarized in the following paragraphs.

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

Where ħ is the reduced planck constant 1.055 x 10^{-34} Js, V(x) represents the potential energy at x, and E represents the summation of both kinetic and potential energies.

1.3.1.1 Semi-empirical methods

These methods are currently utilized for large molecular systems such as biomolecules and protein to render results in an acceptable time frame.[4, 29, 35, 36] Semi-empirical methods are based on quantum mechanics but only valence electrons are considered as a result of severe approximations employed in the method. The common theoretical models in semi-empirical methods are; PM3,

PM6, AM1, INDO and ZINDO.[29, 30, 37, 38] These different theoretical approximations are for specific series of atoms/molecules and their parametrizations include approximations from experimental data or high level of *ab initio* calculation results.[39] Therefore, it is important to verify the suitability of a chosen method for specific chemical systems.

1.3.1.2 *Ab initio* methods

This method remains the most prevalent electronic structure method applied by both computational and theoretical chemist despite the demanding requirements in terms of computational resources and time.[30, 40, 41] *Ab initio* methods are centred on quantum mechanics laws derived from theoretical principles and even though the methods grouped under this model have similar approaches, they vary in the implemented mathematically algorithm approximations.[30, 35, 40, 41] Molecular systems can be studied easily and the required time for such calculation depends on the computer hard and software with respect to the level of theory and basis set applied. *Ab initio* methods includes Configuration Interaction (CI), Hartree-Fock (HF) and Moller-Plesset (MP*n*-including electron correlation).[39, 40] These methods systematically approach accuracy as the level of theory[39, 40] and basis set[39, 40] increases, which comes at a considerable computational cost in terms of time and resources.[30] Reviews on *ab initio* methods have focused on specific studies or calculations as this approach is extensive.[42-45]

1.3.1.3 Basis sets

Basis sets are used to describe electronic wave functions mathematically,[30] molecules are defined by molecular orbitals which are expressed in terms of atomic orbitals.[46] Hence, the larger the basis set the more accurate the computational output. A standard renown basis set for calculating organic molecule properties[47, 48] is the 6-31+G(d,p)[49, 50] and this was applied in three parts of this present study combined with the selected DFT methods. Other examples of basis sets are: DGDZVP,[51, 52] def2-TZVP,[53, 54] MidiX,[55, 56] LANL2DZ,[57, 58] and aug-ccpVTZ[59-61]. The performance, relevance and quality of the these basis sets has been studied[62] and a comprehensive update by Frank[47] and Grant[48] on the various basis sets can be found in literature.

1.3.1.4 Density functional theory methods (DFT)

An alternative method was designed by Hohenberg, Kohn and Sham that uses density functional models for a more effective electronic structural method which computes energies using electron

density rather than wave function.[63, 64] The DFT method can optimize molecular systems up to 3000 atoms using large resources,[65, 66] although the DFT and MP methods produces similar quality results, DFT theory is more cost effective (faster).[66] Time-Dependent DFT is also an extension of DFT method and is extensively used to simulate optical properties of both organic and inorganic compounds,[67] Salahub *et al.*[68] and Laurent *et al.*,[69] presented an update on this model in 2013.

The DFT functionals are basically categorized as: PBE correlation, Dispersion,[70] Becke one-parameter hybrid, Becke three-parameter hybrid,[71] revised B97, Minnesota and ꞇ-dependent gradient-corrected correlational functional, these is presented in Figure 1. The hybrid method Becke3LYP by Becke[72] and Lee *et al.*,[73] is the most common among the DFT models particularly for organic molecules. This model will be applied within the framework of this study using the Gaussian program.[74] This model calculates correlation energies from electron density using exact exchange and gradient corrected density functional approximations.[72, 73, 75] Although, poor estimation of barrier heights and weak exchange-correlations interactions has been noted,[76-78] B3LYP has shown to give relatively good geometries that agree with high level *ab initio* results of organic, molecular and organometallic compounds.[9, 79-83] The M06 functionals are also now widely in use for calculating molecular properties and system modelling with a level of accuracy comparable to experimental results.[62, 84] DFT methods are more attractive because they include electron correlation effects and better results are provided at a comparable cost.[30]

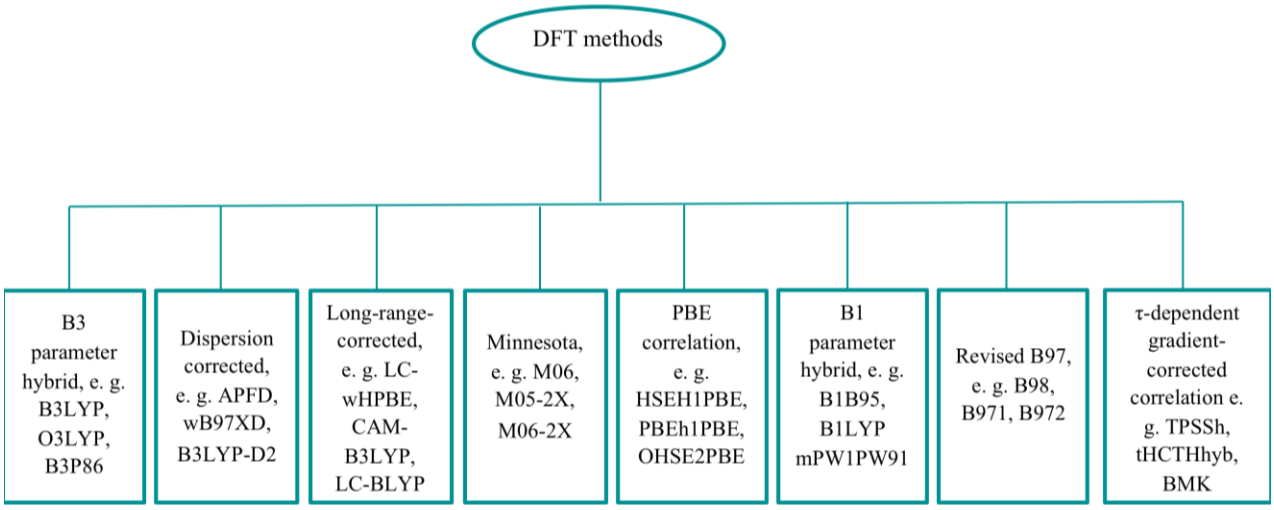


Figure 6. Illustration of the DFT functionals with examples. [26]

1.3.2 Molecular mechanics methods (MM)

This method also known as non-quantum mechanical method uses classical physic laws to understand and predict structural behaviour and properties of a molecule or systems from a single geometry.[32, 85] Large systems such as enzymes and other biological molecules can be modelled with MM methods, and this has an edge over quantum mechanics in terms of computational time and resources as they are characterized by force fields with specific empirical parameters.[30, 86] Universal Force Field (UFF),[87] Merk Molecular Force Field (MMFF)[88] and AMBER[89] are all examples of the molecular mechanics force fields.

For this study, AMBER force field in the Gaussian 16 package[74] was applied which uses a simple harmonic model that includes bond stretch, angle bends, torsion, electrostatic interactions and standard functions for van der Waals.[86, 90, 91] While the Gaussian package assigns the atom types automatically for UFF calculations, atom types needs to be clearly defined for AMBER calculations within the molecule specification section.[30]

Though MM methods cannot be applied to study the bond formation/breaking reaction since it does not consider the electronic effect of a chemical system.[30, 85] However, hybrid QM/MM methods are now being utilized in practical applications, such as; investigating enzymatic processes, the design of drug leads and catalysis, predicting drug metabolism and resistance.[4, 92, 93]

1.4 QM/MM hybrid methods

To overcome the challenges in theoretical modelling of large systems such as proteins/enzymes and considering cost effectiveness, combining the QM with MM method was proposed.[94-96] An extensive review[96] was done on the principles and applications of these methods. The basic concept of the hybrid QM/MM is to split a large system into either two or three parts as regard to research interest(s). The QM part is associated with the reactive atoms that are involved in the catalysis and binding, bond forming/breaking process, while the MM region takes up the rest of the system and solvent.[65, 94]

There are two forms of Hamiltonian QM/MM based on the partitioning strategy, and are known as the additive and subtractive forms,[96] an example is [97, 98] Our own N-layered molecular Orbital

and molecular Mechanics (ONIOM)[65, 99] method. The combined QM/MM approach gives a unique approximation in which the QM accuracy and MM efficiency are both shared in molecular simulations.[96] The QM/MM method was extensively used in this study to unravel the catalytic mechanism of both subtype B and C-SA HIV-1 PR.

1.4.1 ONIOM technique

The ONIOM QM/MM process allows the partitioning of an enzyme or molecular system into layers depending on the relevance of each part to the enzyme's bioactivity and research aim, these parts are named; the model, the intermediate (in the case of a third layer) and the real/environment system.[65] Using the Gaussian software[74] and GaussView,[100] the two- or three-layer segmentation of an entire system can be achieved. For the present investigation, the HIV-1 PR—substrate was partitioned into two distinct layers. The "model" part considers the catalytic site of the enzyme as well as the natural substrate and are treated at the QM; DFT level of theory, whereas the "real" portion consists of the remaining residues of the system and are treated with MM (AMBER forcefield[101]) (Figure 2). The two and three-layer ONIOM calculations are performed in the same manner and are represented respectively as;

$$G_{ONIOM} = G_{real}^{low} + G_{SM}^{high} - G_{SM}^{low} \quad 2$$

$$G_{ONIOM} = G_{real}^{low} + G_{IM}^{medium} + G_{SM}^{high} - G_{IM}^{low} - G_{SM}^{medium} \quad 3$$

This approach has been used to investigate the binding activities of known HIV protease inhibitors with two different HIV-1 PR subtypes and it displays a reasonable level of accuracy comparable with experimental data.[24] Likewise, a review on the ONIOM method with detailed parameterization technique and extensive applications to different classes of molecules and systems is available in literature.[65]

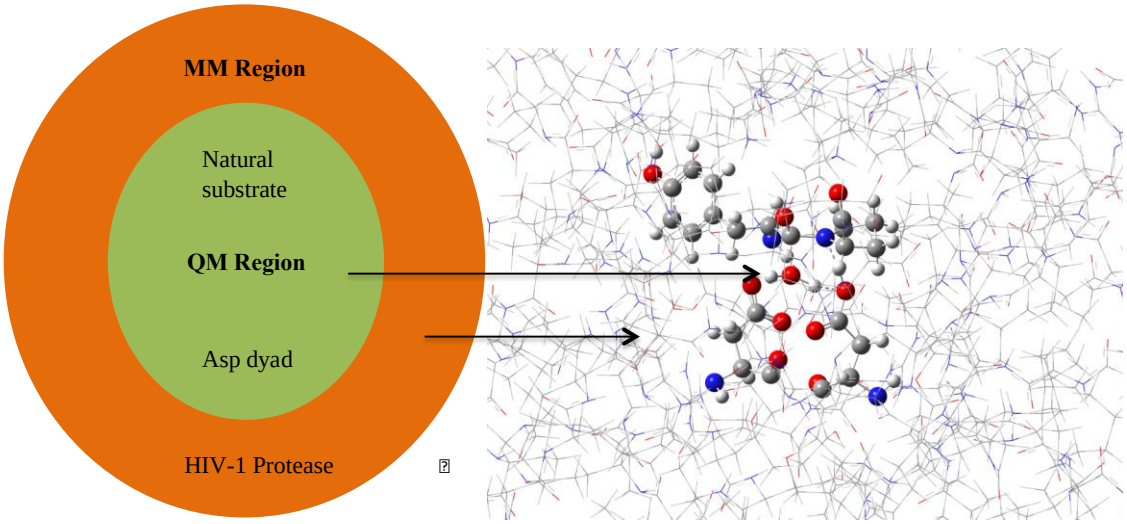


Figure 2. Schematic representation of the two-layered ONIOM model (DFT: AMBER) of HIV-1 PR—natural substrate complex.

1.4.2 Umbrella sampling

This technique in computational chemistry was first developed by Torrie and Valleau in 1977[102] is used to improve sampling of system(s) along a reaction coordinate from one thermodynamic state (reactant) to another (transition state or products).[103] This can be described as a type of combined QM/MM MD method.[102] Basically, an artificial restraint is added to sample some coordinates of a system in a range of values to bias it.[98, 104] These potential biasing minima is moved to another range of coordinates of interest and more simulations are carried out, the various simulations (also known as windows) should have sufficient overlap. The bias potential effect can then be removed, and the potential mean force (PMF) is constructed which is the free energy profile along the selected coordinate.[98, 104] This approach has been widely used to calculate free energy differences in complex systems with AMBER[105] and in the context of this work, this method will be adopted in computing the free energy difference of the natural substrates cleaved by HIV-1 subtypes C-SA.

1.5 Computational software

The computational software required for modelling, calculations and post analyses studies are discussed in the following sections.

1.5.1 Graphical user interfaces (GUIs)

GaussView,[100] Visual Molecular Dynamics (VMD),[106] Discover studio,[107] UCSF Chimera[108] and PyMOL[109] are the pre- and post-processor GUIs program used in this present study. All molecular complexes were modelled and viewed by means of these computational application packages.

1.5.2 The Gaussian program

Gaussian 16 Rev A.03[110] and B.02[74] (2016) are both an upgrade versions of the Gaussian program which started with Gaussian 70,92,94 and 03 versions.[111] Accurate computational methods demand high computer resources, hence electronic structure calculations are usually carried out *via* shared processors. The enzymatic interactions of HIV-1 PR (subtype and C-SA) with its natural substrates was investigated using Gaussian 16 Rev A.03[110] that is installed on clusters at the Centre for High Performance Computing in Cape Town, South Africa (CHPC). This package was used for all calculations including DFT and ONIOM.

1.5.3 The AMBER suite

The AMBER package[112] is one of the common software for conducting MD simulations amongst the software package available.[113] The AMBER 18[105] which is the latest version (2018) has its root from AMBER 4 which represents 12 years of continuous development from version 9 by multiple research groups.[113] AMBER program suite encompasses of numerous other programs working together to setup, perform and analyse MD simulations such as Umbrella sampling.[104] The parameterization of this software package allows the use of graphics processing units (GPUs) that offers more memory bandwidth and computational power which saves computational time and cost.[105] Hence, the AMBER 18[105] suite extension in CHPC (www.chpc.ac.za) was used in our MD simulations and Umbrella sampling for investigating the enzymatic interactions of HIV-1 PR (subtype and C-SA) with its natural substrates.

References

1. Wlodawer, A., and Vondrasek, J. (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design 1, *Annual Review of Biophysics and Biomolecular Structure 27*, 249-284.
2. De Clercq, E. (2004) Antivirals and antiviral strategies, *Nature Reviews Microbiology 2*, 704-720.
3. Wlodawer, A. (2002) Rational approach to AIDS drug design through structural biology*, *Annual Review of Medicine 53*, 595-614.

4. Honarparvar, B., Govender, T., Maguire, G. E. M, Soliman, M. E., and Kruger, H. G. (2013) Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity, *Chemical Reviews 114*, 493-537.
5. Jhoti, H., and Leach, A. R. (2007) Structure-based drug discovery, p. 4, *Springer*.
6. Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014) Computational methods in drug discovery, *Pharmacological Reviews 66*, 334-395.
7. Kapetanovic, I. (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach, *Chemico-biological Interactions 171*, 165-176.
8. Veselovsky, A., and Ivanov, A. (2003) Strategy of computer-aided drug design, *Current Drug Targets-Infectious Disorders 3*, 33-40.
9. Thishana, S., Krishna, B., and Hendrik, G. K. (2006) A computational study of the mechanism of formation of the penta-cycloundecane (PCU) Cage Lactam., *Trends and Perspectives in Modern Computational Science, 7*, 511-514.
10. Schreiner, P. R., Allen, W., Orozco, M., Thiel, W., and Willett, P. (2014) Computational molecular science, *Wiley*.
11. Portela, C., Afonso, C. M., Pinto, M. M., and Ramos, M. J. (2003) Computational studies of new potential antimalarial compounds–stereoelectronic complementarity with the receptor, *Journal of Computer-aided Molecular Design 17*, 583-595.
12. Lacaze-Dufaure, C., Najjar, F., and André-Barrès, C. (2010) First computational evidence of a competitive stepwise and concerted mechanism for the reduction of antimalarial endoperoxides, *The Journal of Physical Chemistry B 114*, 9848-9853.
13. Mammino, L., and Bilonda, M. K. (2014) Computational study of antimalarial pyrazole alkaloids from newbouldia laevis, *Journal of Molecular Modeling 20*, 2464.
14. He, X., Mei, Y., Xiang, Y., Zhang, D. W., and Zhang, J. Z. (2005) Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations, *Proteins: Structure, Function, and Bioinformatics 61*, 423-432.
15. Lee, W., and Engels, B. (2013) Clarification on the decarboxylation mechanism in KasA based on the protonation state of key residues in the acyl-enzyme state, *The Journal of Physical Chemistry B 117*, 8095-8104.
16. Gupta, G., Garci, A., Murray, B. S., Dyson, P. J., Fabre, G., Trouillas, P., Giannini, F., Furrer, J., Süss-Fink, G., and Therrien, B. (2013) Synthesis, molecular structure, computational study and in vitro anticancer activity of dinuclear thiolato-bridged pentamethylcyclopentadienyl Rh (III) and Ir (III) complexes, *Dalton Transactions 42*, 15457-15463.
17. Cui, F., Qin, L., Zhang, G., Liu, Q., Yao, X., and Lei, B. (2008) Interaction of anthracycline disaccharide with human serum albumin: investigation by fluorescence spectroscopic technique and modeling studies, *Journal of Pharmaceutical and Biomedical Analysis 48*, 1029-1036.
18. Bradáč, O., Zimmermann, T., and Burda, J. V. (2013) Can satraplatin be hydrated before the reduction process occurs? the DFT computational study, *Journal of Molecular Modeling 19*, 4669-4680.
19. Mariappan, G., and Sundaraganesan, N. (2014) Spectral and structural studies of the anti-cancer drug flutamide by density functional theoretical method, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 117*, 604-613.

20. Chi, Z., Liu, R., Yang, H., Shen, H., and Wang, J. (2011) Binding of tetracycline and chlortetracycline to the enzyme trypsin: spectroscopic and molecular modeling investigations, *PloS One 6*, e28361.
21. Antony, J., Gresh, N., Olsen, L., Hemmingsen, L., Schofield, C. J., and Bauer, R. (2002) Binding of D-and L-captopril inhibitors to metallo-β-lactamase studied by polarizable molecular mechanics and quantum mechanics, *Journal of Computational Chemistry 23*, 1281-1296.
22. Meliá, C., Ferrer, S., Moliner, V., Tuñón, I., and Bertrán, J. (2012) Computational study on hydrolysis of cefotaxime in gas phase and in aqueous solution, *Journal of Computational Chemistry 33*, 1948-1959.
23. Sanusi, Z. K., Govender, T., Maguire, G. E. M, Maseko, S. B., Lin, J., Kruger, H. G., and Honarparvar, B. (2018) An insight to the molecular interactions of the FDA approved HIV PR drugs against L38L↑ N↑ L PR mutant, *Journal of Computer-aided Molecular Design 32*, 459-471.
24. Sanusi, Z., Govender, T., Maguire, G. E. M, Maseko, S., Lin, J., Kruger, H., and Honarparvar, B. (2017) Investigation of the binding free energies of FDA approved drugs against subtype B and C-SA HIV PR: ONIOM approach, *Journal of Molecular Graphics and Modelling 76*, 77-85.
25. Makatini, M. M., Petzold, K., Arvidsson, P. I., Honarparvar, B., Govender, T., Maguire, G. E. M, Parboosing, R., Sayed, Y., Soliman, M. E., and Kruger, H. G. (2012) Synthesis, screening and computational investigation of pentacycloundecane-peptoids as potent CSA-HIV PR inhibitors, *European Journal of Medicinal Chemistry 57*, 459-467.
26. Makatini, M. M., Petzold, K., Alves, C. N., Arvidsson, P. I., Honarparvar, B., Govender, P., Govender, T., Kruger, H. G., Sayed, Y., and JerônimoLameira. (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors, *Journal of Enzyme Inhibition and Medicinal Chemistry 28*, 78-88.
27. Honarparvar, B., Pawar, S. A., Alves, C. N., Lameira, J., Maguire, G. E. M, Silva, J. R. A., Govender, T., and Kruger, H. G. (2015) Pentacycloundecane lactam vs lactone norstatine type protease HIV inhibitors: binding energy calculations and DFT study, *Journal of Biomedical Science 22*, 1-15.
28. Hehre, W. J. (2003) A guide to molecular mechanics and quantum chemical calculations, Vol. 2, *Wavefunction Irvine, CA*.
29. Schreiner, P. R. (2007) Relative energy computations with approximate density functional theory—a caveat!, *Angewandte Chemie International Edition 46*, 4217-4219.
30. Foresman, J., and Frish, E. (1996) Exploring chemistry, *Gaussian Inc., Pittsburg, USA*.
31. Schrödinger, E. (1926) An undulatory theory of the mechanics of atoms and molecules, *Physical Review 28(6)*, 1049.
32. Atkins, P., and De Paula, J. (2006) Atkins' physical chemistry, *New York*.
33. Joseph-McCarthy, D. (1999) Computational approaches to structure-based ligand design, *Pharmacology & Therapeutics 84*, 179-191.
34. Cramer, C. J. (2013) Essentials of computational chemistry: theories and models, *John Wiley & Sons*.
35. Lewars, E. G. (2010) Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics, *Springer Science & Business Media*.

36. Nicolaides, C. A. (2014) Quantum chemistry and its "ages", *International Journal of Quantum Chemistry 114*, 963-982.
37. Reimers, J. R. (2011) Computational methods for large systems: electronic structure approaches for biotechnology and nanotechnology, *John Wiley & Sons*.
38. Liu, X.-T., Zhao, Y., Ren, A.-M., and Feng, J.-K. (2011) A comparative study of one-and two-photon absorption properties of pyrene and perylene diimide derivatives, *Journal of Molecular Modeling 17*, 1413-1425.
39. Frisch, A., and Foresman, J. (1996) Exploring chemistry with electronic structure methods, *Pittsburgh PA: Gaussian Inc 302*.
40. Ghosh, S. K., and Chattaraj, P. K. (2013) Concepts and Methods in Modern Theoretical Chemistry: Electronic Structure and Reactivity, *CRC Press*.
41. Leszczynski, J. (2012) Handbook of computational chemistry, Vol. 2, *Springer Science & Business Media*.
42. Jiang, G., Chen, W., and Zheng, Y. (2017) A review of recent ab initio studies on strain-tunable conductivity in tunnel junctions with piezoelectric, ferroelectric and multiferroic barriers, *Semiconductor Science and Technology 32(8)*, 083006.
43. Brommer, P., Kiselev, A., Schopf, D., Beck, P., Roth, J., and Trebin, H.-R. (2015) Classical interaction potentials for diverse materials from ab initio data: a review of potfit, *Modelling and Simulation in Materials Science and Engineering 23(7)*, 074002.
44. Ishikawa, K. L., and Sato, T. (2015) A review on ab initio approaches for multielectron dynamics, *IEEE Journal of Selected Topics in Quantum Electronics 21*, 1-16.
45. Yanai, T., Kurashige, Y., Mizukami, W., Chalupský, J., Lan, T. N., and Saitow, M. (2015) Density matrix renormalization group for ab initio calculations and associated dynamic correlation methods: a review of theory and applications, *International Journal of Quantum Chemistry 115*, 283-299.
46. Hinchliffe, A. (2003) Molecular modelling for beginners, *John Wiley & Sons Ltd, Chichester*.
47. Jensen, F. (2013) Atomic orbital basis sets, *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 273-295.
48. Hill, J. G. (2013) Gaussian basis sets for molecular applications, *International Journal of Quantum Chemistry 113*, 21-34.
49. Hehre, W. J., Ditchfield, R., and Pople, J. A. (1972) Self—consistent molecular orbital methods. XII. Further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules, *The Journal of Chemical Physics 56*, 2257-2261.
50. Pople, J. A., Gill, P. M., and Johnson, B. G. (1992) Kohn—Sham density-functional theory within a finite basis set, *Chemical Physics Letters 199*, 557-560.
51. Sosa, C., Andzelm, J., Elkin, B. C., Wimmer, E., Dobbs, K. D., and Dixon, D. A. (1992) A local density functional study of the structure and vibrational frequencies of molecular transition-metal compounds, *The Journal of Physical Chemistry 96*, 6630-6636.
52. Godbout, N., Salahub, D. R., Andzelm, J., and Wimmer, E. (1992) Optimization of Gaussian-type basis sets for local spin density functional calculations. part I. boron through neon, optimization technique and validation, *Canadian Journal of Chemistry 70*, 560-571.
53. Weigend, F., and Ahlrichs, R. (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy, *Physical Chemistry Chemical Physics 7*, 3297-3305.

54. Weigend, F. (2006) Accurate Coulomb-fitting basis sets for H to Rn, *Physical Chemistry Chemical Physics 8*, 1057-1065.
55. Li, J., Cramer, C. J., and Truhlar, D. G. (1998) MIDI! basis set for silicon, bromine, and iodine, *Theoretical Chemistry Accounts 99*, 192-196.
56. Easton, R. E., Giesen, D. J., Welch, A., Cramer, C. J., and Truhlar, D. G. (1996) The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges, *Theoretica Chimica Acta 93*, 281-301.
57. Hay, P. J., and Wadt, W. R. (1985) Ab initio effective core potentials for molecular calculations. potentials for K to Au including the outermost core orbitals, *The Journal of Chemical Physics 82*, 299-310.
58. Wadt, W. R., and Hay, P. J. (1985) Ab initio effective core potentials for molecular calculations. potentials for main group elements Na to Bi, *The Journal of Chemical Physics 82*, 284-298.
59. Dunning Jr, T. H. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *The Journal of Chemical Physics 90*, 1007-1023.
60. Kendall, R. A., Dunning Jr, T. H., and Harrison, R. J. (1992) Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions, *The Journal of Chemical Physics 96*, 6796-6806.
61. Woon, D. E., and Dunning Jr, T. H. (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon, *The Journal of Chemical Physics 98*, 1358-1371.
62. Lawal, M. M., Govender, T., Maguire, G. E. M, Honarparvar, B., and Kruger, H. G. (2016) Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study, *Journal of Molecular Modeling 22*, 235-246.
63. Hohenberg, P., and Kohn, W. (1964) Inhomogeneous electron gas, *Physical Review 136*, B864.
64. Kohn, W., Becke, A. D., and Parr, R. G. (1996) Density functional theory of electronic structure, *The Journal of Physical Chemistry 100*, 12974-12980.
65. Chung, L. W., Sameera, W., Ramozzi, R., Page, A. J., Hatanaka, M., Petrova, G. P., Harris, T. V., Li, X., Ke, Z., and Liu, F. (2015) The ONIOM method and its applications, *Chemical Reviews 115*, 5678-5796.
66. Görling, A. (1992) Kohn-Sham potentials and wave functions from electron densities, *Physical Review A 46*, 3753.
67. Tomberg, A. (2013) Gaussian 09W Tutorial, An introduction to computational chemistry using G09W and avogadro software, 1-36.
68. Salahub, D. R., De La Lande, A., Goursot, A., Zhang, R., and Zhang, Y. (2013) Recent progress in density functional methodology for biomolecular modeling, in applications of density functional theory to biological and bioinorganic chemistry, pp 1-64, *Springer*.
69. Laurent, A. D., and Jacquemin, D. (2013) TD-DFT benchmarks: a review, *International Journal of Quantum Chemistry 113*, 2019-2039.
70. Kashinski, D., Chase, G., Nelson, R., Di Nallo, O., Scales, A., VanderLey, D., and Byrd, E. (2017) Harmonic vibrational frequencies: approximate global scaling factors for TPSS, M06, and M11 functional families using several common basis sets, *The Journal of Physical Chemistry A 121*, 2265-2273.

71. Wang, M. Y.-W. (2018) The Application of genetic algorithms for density functional optimization and development, *Middle Tennessee State University*.
72. Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange, *The Journal of Chemical Physics 98*, 5648-5652.
73. Lee, C., Yang, W., and Parr, R. G. (1988) Development of the colle-salvetti correlation-energy formula into a functional of the electron density, *Physical Review B 37*, 785-789.
74. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Petersson, G., and Nakatsuji, H. Gaussian 16 Revision B. 01. 2016; Gaussian Inc, *Wallingford CT*.
75. Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A 38(6)*, 3098.
76. Zhao, Y., González-García, N., and Truhlar, D. G. (2005) Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods, *The Journal of Physical Chemistry A 109*, 2012-2018.
77. Zhao, Y., and Truhlar, D. G. (2008) Density functionals with broad applicability in chemistry, *Accounts of Chemical Research 41*, 157-167.
78. Chéron, N., Jacquemin, D., and Fleurat-Lessard, P. (2012) A qualitative failure of B3LYP for textbook organic reactions, *Physical Chemistry Chemical Physics 14*, 7170-7175.
79. Singh, T., Kruger, H. G., Bisetty, K., and Power, T. D. (2012) Theoretical study on the formation of a pentacyclo-undecane cage lactam, *Computational and Theoretical Chemistry 986*, 63-70.
80. Gokul, V., Kruger, H. G., Govender, T., Fourie, L., and Power, T. D. (2004) An ab initio mechanistic understanding of the regioselective acetylation of 8, 11-dihydroxy-pentacyclo [5.4. 0.02, 6.03, 10.05, 9] undecane-8, 11-lactam, *Journal of Molecular Structure: Theochem 672*, 119-125.
81. Kruger, H. G., Mdluli, P., Power, T. D., Raasch, T., and Singh, A. (2006) Experimental and computational studies of the regioselective protection of hydantoins using anhydride, *Journal of Molecular Structure: THEOCHEM 771*, 165-170.
82. Ansbacher, T., Srivastava, H. K., Martin, J. M., and Shurki, A. (2010) Can DFT methods correctly and efficiently predict the coordination number of copper (I) complexes? a case study, *Journal of Computational Chemistry 31*, 75-83.
83. Hadebe, S. W., Kruger, H. G., and Robinson, R. S. (2011) A DFT study of the hydroboration reaction with oxygen-, sulphur-, and nitrogen-based boranes, *Computational and Theoretical Chemistry 968*, 26-30.
84. Lawal, M. M., Govender, T., Maguire, G. E. M, Kruger, H. G., and Honarparvar, B. (2018) DFT study of the acid-catalyzed esterification reaction mechanism of methanol with carboxylic acid and its halide derivatives, *International Journal of Quantum Chemistry 118*, 25497-25508.
85. Bultinck, P., De Winter, H., Langenaeker, W., and Tollenare, J. P. (2003) Computational Medicinal Chemistry for Drug Discovery, *CRC Press*.
86. Rogers, D. W. (2003) Computational chemistry using the PC, *John Wiley & Sons*.
87. Rappé, A. K., Casewit, C. J., Colwell, K., Goddard Iii, W., and Skiff, W. (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *Journal of the American Chemical Society 114*, 10024-10035.

88. Halgren, T. A. (1996) Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94, *Journal of Computational Chemistry 17*, 490-519.
89. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.
90. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins, *Journal of the American Chemical Society 106*, 765-784.
91. Sansom, C. E., and Smith, C. A. (1998) Computer applications in the biomolecular sciences. part 1: molecular modelling, *Biochemical Education 26*, 103-110.
92. De Benedetti, P. G., and Fanelli, F. (2010) Computational quantum chemistry and adaptive ligand modeling in mechanistic QSAR, *Drug Discovery Today 15*, 859-866.
93. Verma, J., Khedkar, V. M., and Coutinho, E. C. (2010) 3D-QSAR in drug design-a review, *Current Topics in Medicinal Chemistry 10*, 95-115.
94. Warshel, A., and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *Journal of Molecular Biology 103*, 227-249.
95. Warshel, A. (1991) Computer modelling of chemical reactions in enzymes and solutions.
96. Xu, D., Zheng, M., and Wu, S. (2012) Principles and applications of hybrid quantum mechanical and molecular mechanical methods, in quantum simulations of materials and biological systems, pp 155-168, *Springer*.
97. Torrie, G. M., and Valleau, J. P. (1974) Monte carlo free energy estimates using non-boltzmann sampling: application to the sub-critical lennard-jones fluid, *Chemical Physics Letters 28*, 578-581.
98. Kästner, J. (2011) Umbrella sampling, *Wiley Interdisciplinary Reviews: Computational Molecular Science 1*, 932-942.
99. Svensson, M., Humbel, S., Froese, R. D., Matsubara, T., Sieber, S., and Morokuma, K. (1996) ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition, *The Journal of Physical Chemistry 100*, 19357-19363.
100. Dennington, R., Keith, T., and Millam, J. (2009) Semichem Inc, *Shawnee Mission KS, GaussView, Version 5*.
101. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., and Lee, T. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *Journal of Computational Chemistry 24*, 1999-2012.
102. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method, *Journal of Computational Chemistry 13*, 1011-1021.
103. Hazel, A., and Gumbart, J. C. (2017) Methods for calculating potentials of mean force.
104. Case, D., Cerutti, D., Cheatham III, T., Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., Greene, D., and Homeyer, N. (2018) AMBER reference manual, *University of California*.

105. Case, D., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham III, T., Cruzeiro, V., Darden, T., Duke, R., Ghoreishi, D., and Gilson, M. AMBER 2018; 2018, *University of California, San Francisco*.
106. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *Journal of Molecular Graphics 14*, 33-38.
107. BIOVIA, D. S., and PipelinePilot, R. Dassault Systèmes; San Diego: 2015, Discovery Studio Modeling Environment, *Release 4*.
108. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.
109. DeLano, W. L. (2002) The PyMOL molecular graphics system.
110. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Petersson, G., and Nakatsuji, H. (2016) Gaussian 16, Revision A. 03, Gaussian, *Inc., Wallingford CT*.
111. GWT, M. F., Schlegel, H., Scuseria, G., Robb, M., and Cheeseman, J. (2003) Gaussian 03, *Revision B 1*.
112. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, *Computer Physics Communications 91*, 1-41.
113. Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013) An overview of the amber biomolecular simulation package, *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 198-210.

FINAL

Send Orders for Reprints to reprints@benthamscience.net

Current Medicinal Chemistry, 2019, *26*, 1-34

REVIEW ARTICLE

From Recognition to Reaction Mechanism: An Overview on the Interactions between HIV-1 Protease and its Natural Targets

Monsurat M. Lawal[a], Zainab K. Sanusi[a], Thavendran Govender[a], Glenn E.M. Maguire[a,b], Bahareh Honarparvar[a], and Hendrik G. Kruger[a,*]

[a]*Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa;* [b]*School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa*

ARTICLE HISTORY

Received: May 23, 2018
Revised: November 04, 2018
Accepted: November 07, 2018

DOI:
10.2174/0929867325666181113122900

Abstract: Current investigations on the human immunodeficiency virus protease (HIV-1 PR) as a druggable target towards the treatment of AIDS require an update to facilitate further development of promising inhibitors with improved inhibitory activities. For the past two decades, up to 100 scholarly reports appeared annually on the inhibition and catalytic mechanism of HIV-1 PR. A fundamental literature review on the prerequisite of HIV-1 PR action leading to the release of the infectious virion is absent. Herein, recent advances (both computationally and experimentally) on the recognition mode and reaction mechanism of HIV-1 PR involving its natural targets are provided. This review features more than 80 articles from reputable journals. Recognition of the natural Gag and Gag-Pol cleavage junctions by this enzyme and its mutant analogs was first addressed. Thereafter, a comprehensive dissect of the enzymatic mechanism of HIV-1 PR on its natural polypeptide sequences from literature was put together. In addition, we highlighted ongoing research topics in which *in silico* methods could be harnessed to provide deeper insights into the catalytic mechanism of the HIV-1 protease in the presence of its natural substrates at the molecular level. Understanding the recognition and catalytic mechanism of HIV-1 PR leading to the release of an infective virion, which advertently affects the immune system, will assist in designing mechanism-based inhibitors with improved bioactivity.

Keywords: HIV-1 PR, natural substrates, recognition pattern, reaction mechanism, transition state modeling.

1. INTRODUCTION

Aspartate proteases (Asp PRs) form a distinct class of hydrolytic enzymes with two characteristic aspartate residues acting as the major catalyst in their active sites, and these Asp PRs cleave specific proteins or polypeptides [1-3]. Notable amongst this class of enzymes are beta-secretase 1 (BACE 1; EC 3.4.23.46), [4] cathepsin D (Cat D; EC 3.4.23.5), [5-7] human immunodeficiency virus type 1 protease (HIV-1 PR; EC 3.4.23.16) [8] and plasmepsin II (Plm II; EC 3.4.23.39) [9]. These proteinases are currently receiving extensive attention as potential drug targets in a number of serious infections and diseases, which include Alzheimer's disease (AD), acquired Immunodeficiency Syndrome (AIDS) and malaria [10]. HIV-1 PR is an indispensable drug target moiety towards the treatment of HIV/AIDS [11].

HIV-1 PR is a typical lentivirus proteinase, which is homodimeric with normally a total number of 99 amino acid residues in each monomer. However, variants of 100 (I36T↑T) [12] and 101 (L38L↑N↑L) [13] amino acid residues were recently reported by our research group. HIV PR is an important degrading enzyme necessary for the proteolytic cleavage of the Gag and Gag-Pol polyproteins, required for the development of mature virion proteins [11, 14]. The hydrolytic action of the PR on these asymmetric natural polyprotein substrate sequences results in the processing of the corre-

*Address correspondence to this author at the Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa. E-mail: kruger@ukzn.ac.za

0929-8673/19 $58.00+.00

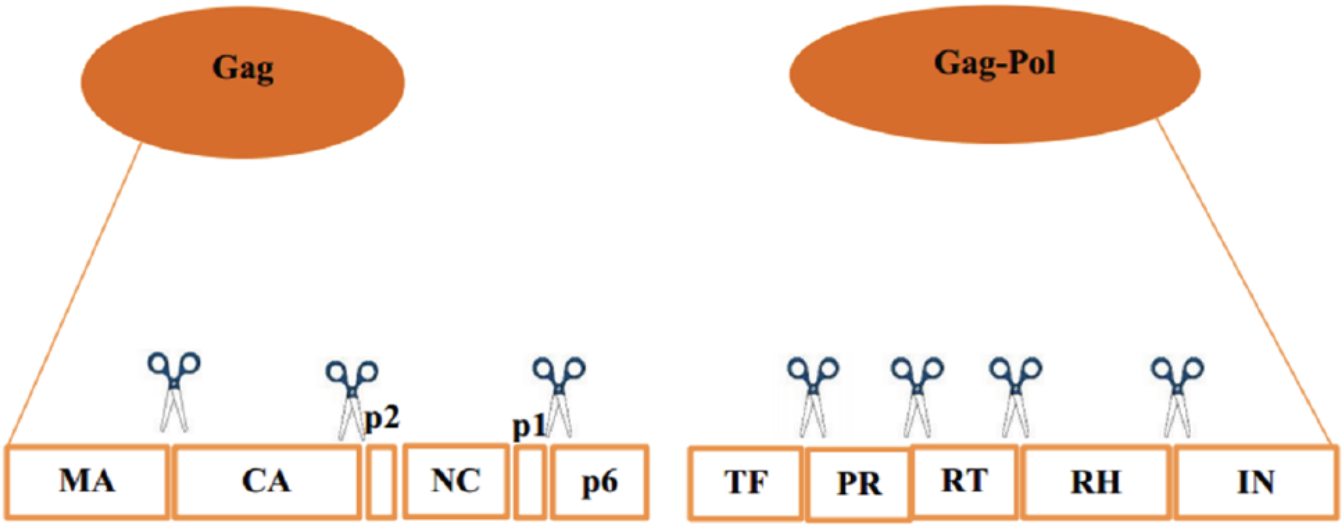


Fig. (1). Illustration of Gag and Gag-Pol cleavage process by HIV-1 PR.

sponding mature virion: Gag proteolysis leads to capsid (CA), matrix (MA), spacer peptide 1 (p1), nucleocapsid (NC), spacer peptide 2 (p2) and p6gag [15]. Gag-Pol leads to reverse transcriptase (RT), RNAse H (RH), and integrase (IN) [16]. The mechanism of action of PR on these polyproteins has been a subject of research over the past three decades and discussed herein. When HIV-1 PR cleaves these sites, nine cleavage domains are produced (Fig. **1** and Table **1**).

In 2000, a three-dimensional (3-D) structure of HIV-1 PR complexed with a natural substrate (corresponding to CA-p2) was deposited by Prabu-Jeyabalan *et al.* [17] in the Protein Data Bank (PDB) with a resolution of 2.0 Å (PDB code 1F7A [18], Fig. **2**). The substrate polypeptide sequence, containing ten amino acid residues, is long enough to cover the whole binding epitope of the protease which makes this result different from the other 19 HIV-1 PR—substrate bound crystal structures deposited in the PDB earlier [17]. Those structures are relatively short covering P3-P3′ epitope and most of which are not a close representation of the actual natural substrate segments [17]. Since the inception of this study [17], about 80 research articles (Web of Science Database, accessed on 29 September 2017) have focused on the natural substrate and HIV-1 PR demonstrating the importance of this topic. The cleavage sites of the HIV-1 PR corresponding to the Gag and Gag-Pol proteins precursor are represented in Table **1**.

Experimental investigation on the interactions between HIV-1 PR and the natural substrate include: how this homodimeric enzyme recognizes asymmetric substrates, [17, 19] natural substrates mimics for the design of potent HIV-1 PR inhibitors, [20, 21] crystallization of HIV-1 PR—natural substrate complexes, [17, 19] mutations at cleavage sites, [22] and the determination of the enzyme processing rate of its substrates [23]. The application of *in silico* methods that better reveal the interaction between natural substrates and the HIV-1 PR including the reaction mechanism have been reported [10, 24-29]. Despite a vast number of such contributions from across the globe, a comprehensive review on experimental as well as theoretical understanding of substrate specificity (in terms of natural target cleavage domains) and the substrate-PR reaction mechanism (Table **1**) is lacking.

A review that covers substrate sequences recognition and catalytic reaction mechanism of HIV-1 PR is required to provide tenable answers to pertinent questions on this perspective. Some of these questions are: what is the recognition pattern of HIV-1 PR—natural substrate system? What are the possible mechanistic models for the catalysis of natural substrates by HIV-1 PR? Can a single protonation pattern be sufficient for the description of all the natural substrates catalysis? What is the most suitable representation of HIV-1 PR—natural substrate complex in computational modeling of the cleavage mechanism? Can the structural moieties along the reaction coordinate (observed in theoretical simulations) be captured through experimental method(s)? Is the mechanistic route of drug-resistant and the native HIV-1 PRs catalysis of substrates the same?

Meanwhile, we found a number of review articles where HIV-1 PR—natural substrate reaction process was highlighted [11, 30-32]. In their 2000 report, Gul-

nik *et al.* [30] wrote a review in which two subsections were dedicated to substrate specificity, cleavage mechanism, as well as protonation state of the catalytic aspartates. In addition to summarizing the enzymatic function of HIV-1 PR, these reviewers [30] provided a vivid description of the enzyme, documented the biochemical and biophysical properties, highlighted the kinetic constants of the natural substrates and discussed mutations arising from HIV-1 PR drug resistance.

Table 1. The nine recognition non-homologous natural substrate polypeptides segments cleaved by the HIV-1 protease [17, 19].

Peptide Sequences Cleavage Domain	Natural Substrate
Cleavage sites in Gag	
Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-Asn	MA-CA
Lys-Ala-Arg-Val-Leu*Ala-Glu-Ala-Met-Ser	CA-p2
Pro-Ala-Thr-Ile-Met*Met-Gln-Arg-Gly-Asn	p2-NC
Glu-Arg-Gln-Ala-Asn*Phe-Leu-Gly-Lys-Ile	NC-p1
Arg-Pro-Gly-Asn-Phe*Leu-Gln-Ser-Arg-Pro	p1-p6
Cleavage sites in Gag-Pol	
Val-Ser-Phe-Asn-Phe*Pro-Gln-Ile-Thr-Leu	TF-PR
Cys-Thr-Leu-Asn-Phe*Pro-Ile-Ser-Pro-Ile	PR-RT
Gly-Ala-Glu-Thr-Phe*Tyr-Val-Asp-Gly-Ala	RT-RH
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	RH-IN

The asterisk (*) denotes the scissile bond. Matrix-capsid; MA-CA, capsid-p2; CA-p2, p2-nucleopsid; p2-NC, nucleopsid-p1; NC-p1, *trans* frame peptide-protease; TF-PR, protease-reverse transcriptase; PR-RT, reverse transcriptase-RNAseH; RT-RH, RNAseH-integrase; RH-IN, auto proteolysis; AutoP [17, 19]

In 2003, Brik and Wong [33] presented a perspective on the HIV-1 PR mechanism and its implications for drug design. The overview spans studies on the mechanism of HIV-1 PR, drug design and PR resistance. With respect to the natural substrate, three plausible catalytic models were enumerated (Schemes **S1–S3**; Supporting Information, **SI**) based on the reviewed literature. The authors proposed the following research questions: [33] where is the exact location of the proton on the carboxyl groups of the catalytic Asps and what is the role of the flaps in the enzymatic reaction of HIV-1 PR?

Rodríguez-Barrios and Gago in 2004 [31] also gave a brief description of this proteolytic enzyme and provided a summary of the HIV-1 PR hydrolytic mechanism. In 2013, *in vitro* and *in silico* studies on the protonation states of the reactive Asp moieties of this homodimeric enzyme were reviewed [34] alongside other aspartyl protease families. This contribution [34] provided details of experimental and computational catalysis with respect to inhibitors, protonation states induced by inhibitors, probing the electronic state *via* pH measurement of the titratable residues. An estimation of the pK_a values for the acidic residues of HIV-1 PR-inhibitor complexes was also reported. It was suggested that when screening potential ligands for this enzyme, protonation assignment on the ionisable Asp dyad should be based on the results of titration curves and protonation state at a given pH [34].

Herein, we provide a detailed account on applicable research (since 2000) on the catalytic reaction mechanisms of HIV-1 PR and its natural substrates; more than 100 references from the literature are cited. This review covers the basis for recognition and possible reaction pathways of the HIV-1 PR complexed with its natural substrates, the last part of the review provides a future perspective concerning this topic.

2. RECOGNITION OF THE ASYMMETRIC SUBSTRATE BY HOMODIMERIC HIV-1 PR

For an enzyme to carry out its bioactivity, molecular recognition of its substrates/inhibitors is imperative. Processing of the natural Gag and Gag-Pol polyproteins by HIV-1 PR is crucial for viral assembly and replication in the HIV life cycle [35]. In this section, we highlight the gained knowledge on the recognition of natural substrates by HIV-1 PR and mutant variants.

2.1. The Natural Substrate

The hydrolytic cleavage of certain segments in the natural Gag and Gag-Pol polyprotein precursors by HIV-1 PR (Fig. **1**) leads to the production of specific polypeptide units, which are subsequently processed, into separated peptide units. Nine cleavage sites of the Gag and Gag-Pol proteins precursor are recognized and cleaved by HIV-1 PR to produce corresponding peptide segments of the natural substrates (Table **1**) [17, 19]. The mechanism of substrate hydrolysis by this PR has remained a major backbone of inhibitor design [36-38] for the treatment of HIV/AIDS [11, 39].

2.1.1. Cleavage Points

Provided in (Fig. **1**), are the cleavage domains within the Gag and Gag-Pol gene which are recognised by the HIV-1 PR. The recognition mechanism and specificity of HIV-1 PR is mostly associated with substrate modulation, conserved substrate shape, interdependence conformational adaptability of both HIV-1

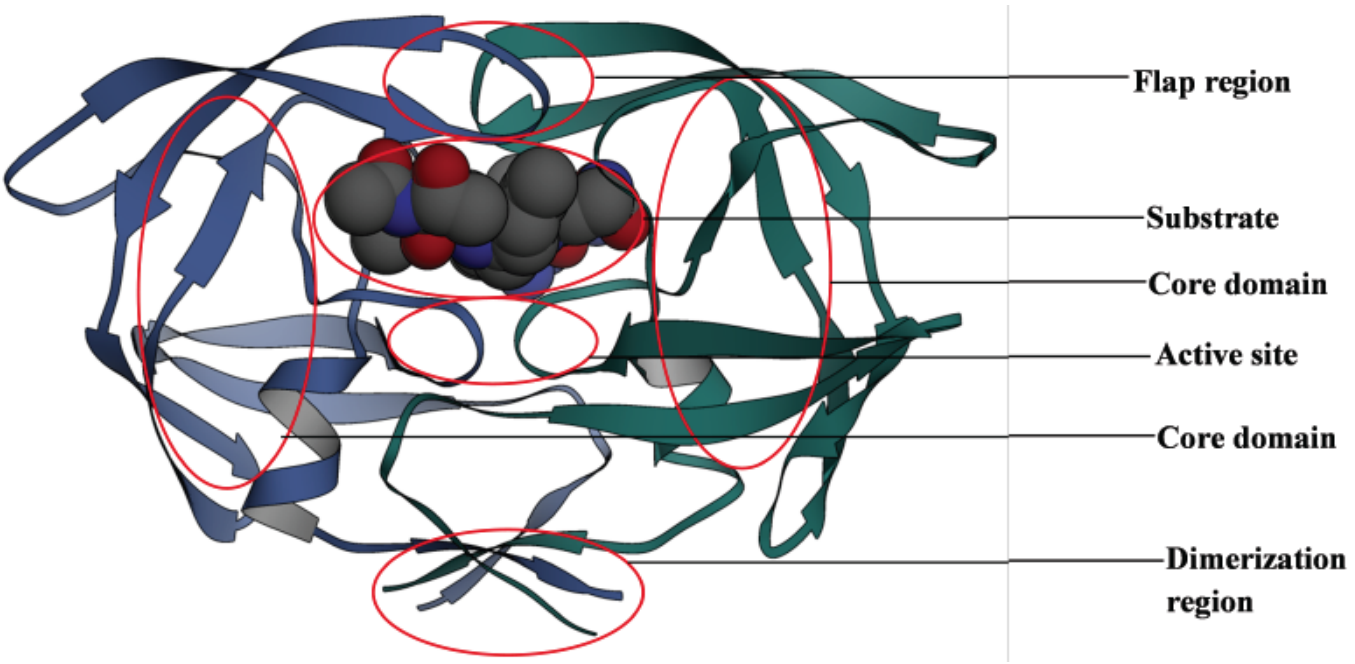


Fig. (2). Structure of HIV-1 protease complexed with CA-p2. PDB code 1F7A[17] (RCSB PDB accessed on 28 March 2017). Image created with UCSF Chimera [30.]

PR and substrates [17, 19, 40-45]. A contrasting opinion is that substrate recognition of HIV-1 PR seems to be based on the conformational specificity of the protease to the Gag and Gag-Pol polyprotein precursors [46]. The non-homologous peptide sequences of each cleavage domain have a characteristic scissile bond (C—N) which is the attacking point for the hydrolytic action of the HIV-1 PR. Mimics of the scissile amino acids within these cleavage domains form the basis of inhibitor design for this enzyme, most importantly the Phe*Pro scissile units (Table **1**) [47].

2.2. Recognition of Natural Substrates by HIV-1 PR

The processing and recognition of CA-p2 cleavage domain within the Gag polypeptide sequence by HIV-1 PR was studied using a crystallographic approach [17]. The authors noted an extensive binding landscape in which the asymmetry peptide forms 24 hydrogen bonds which do not account for substrate recognition. They, however, linked the ability of HIV-1 PR to identify its substrate with the interdependence of the conformational changes. This fluctuation is mediated by six water molecules forming hydrogen bonds that bridge the peptide chains as well as van der Waals (VDW) contact contributions from the protease's amino acid residues [17]. An addendum to this, research provides another intriguing investigation where the same authors [17, 19] determined the X-ray crystal structures of five HIV-1 PR—substrate complexes [19] (deposited in the PDB) [18]. The sequences of the investigated substrates consist of ten asymmetric amino acid units, which cover about 1000 $Å^2$ of the binding surface area. They reiterated that HIV-1 PR recognizes the shape of the non-homologous peptide substrate and not a particular amino acid sequence [17, 19].

Weber and coworkers [48] presented a comparative analysis between human T-cell leukaemia virus (HTLV-1) PR and HIV-1 PR through substrate binding site examination. An oligopeptide containing nine amino acids corresponding to the MA-CA cleavage domain (Table **1**) was studied along with the CA-NC cleavage site of the former enzyme. Combining enzyme assays with molecular modeling methods, they observed a wide difference in the specificity of the HTLV-1 PR and HIV-1 PR. Variance of the substrate sequence at different positions reveals that most retroviral PRs, including HIV-1, preferentially enjoy large hydrophobic side chains at the P1 position. The amino acid in this position is located within the substrate pocket on the S1 subsite and seems to be highly conserved [48]

Light emitting holoprotein was also employed to detect the proteolytic bond cleavage site of the MA-CA precursor [35]. This was revealed by a decrease in the bioluminescence generated by the aequorin fusion protein in the solid phase. The bioluminescence method yielded very sensitive detection limits (1×10^{-11} M) of

the substrate's peptide bond cleavage by the HIV-1 PR [35]. Although the design of the study does not provide substantial input on how the HIV-1 PR recognizes its substrates, it is interesting to also mention that the use of this cysteine-free mutant of aequorin was harnessed to monitor the activities of two competitive and one non-competitive inhibitor for the HIV-1 protease [35].

In 2006, Schiffer and coworkers [42] presented a theoretical approach to evaluate the preferences of substrate positions and correlations between them that might also identify which positions within known substrates can likely tolerate sequence variability and which cannot. This was done using a biased sequence search threading (BSST) method [49]. The prediction of the eight non-identical amino acid sequences of the substrates was investigated using a probability function which is dependent on residue positions within the BSST generated sequence. This was done by employing three models; (1) each substrate position was considered independently, (2) interdependence between pairs was monitored and (3) triple-wise interdependency was tested [42]. Again, it was resolved that HIV-1 PR identifies the substrate shape in its entirety with little or nil cognizance of its specific sequence [42].

A closely related research initiative is an all-atom simulation study by Perez *et al.* [46]. CA-p2 cleavage domain (consisting of eight amino acid sequences) and a non-substrate was studied using molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) and thermodynamic integration (TI) [50-52] to determine substrate detection in HIV-1 PR. The obtained binding free energies were drawn at 20 and 40 ns and the affinity of non-substrate for the PR was found to be high [46]. The authors [46] estimated the free energy (ΔG_{bind}) change between the non-substrate and studied natural substrate to be -4.4 kcal mol^{-1}. Disputing both induced-fit and lock-and-key models as tenable descriptions for substrate recognition, these authors adopted the following proposition to elucidate substrate recognition by HIV-1 PR [46]. (1) The PR preferentially cleaves only natural substrates, with enough exposure, for a given period of time. In other words, even with greater affinity, many non-substrates are not cleaved because they are concealed within the polyprotein pool and apparently unrecognizable by the PR. (2) From the adequately seen sequences, the detection is based on the conformational specificity of PR:Gag and PR:Gag-Pol polyprotein complexes that regulate which residues are within an accessible position to the active site [46].

The distinctive binding scope of HIV-1 PR—substrate complex (referred to as "the substrate envelope") was redefined [53] using molecular dynamics (MD) simulations [54]. This study was a sequel investigation from earlier reported work [40, 55-58] on the design of inhibitors against HIV-1 PR to test the hypothesized substrate envelope. It was proposed that the conserved conformation with a characteristic overlapping cavity in which the substrate sits within the active site, is the substrate envelope [40]. Özen *et al.* [53] carried out MD simulations of seven natural substrates with HIV-1 PR at 11 ns to provide a clearer picture of the conserved 3-D geometry attained during HIV-1 PR—substrate binding. The dynamic substrate envelope was thereafter emphasized to be an exact representation of HIV-1 PR—substrate interactions [53].

Schiffer and coworkers [59] attempted to initiate a model in which an engineered HIV-1 PR (induced mutations) could be used to monitor substrate binding specificity *via* computational techniques. In simple terms, the investigation aimed at employing mathematical, physicochemical, biochemistry and engineering processes to develop an improved computational algorithm that provides a comprehensive description of altered HIV-1 PR substrate recognition. It was observed that substrate specificity was modified in mutated HIV-1 PR, thereby facilitating alternate binding modes for the natural substrates. Rationalizing the weaknesses of the procedure and the output, these authors [59] emphasized that sound knowledge on the structural and conformational plasticity of the enzyme is vital to parameterizing such an algorithm. They proposed that more accurate and advanced computational methods would probably be required to predict the conformational backbones of the remodeled proteins. For side-chain optimization, these authors have used Dunbrack and Cohen-based backbone-dependent rotamer library [60] and the dead-end elimination (DEE) theorem was used for the design of individual pockets [59].

Related topics under this section include, analysis of the context surrounding the processing sites to determine the cleavage rate of Gag and Gag-Pol polyprotein precursors, [23, 61, 62] and identification of efficiently cleaved substrates of HIV-1 PR [63]. It was claimed that identification of substrate cleavable sites could serve as the best templates for the synthesis of the better binding inhibitors [63]. An overall conclusion was also reached that the selection of processing sites and the order of precursor processing are defined partly by the structure of Gag-Pol itself and that this conformation assists in estimating the order of the initial cleav-

age events [62]. Using fluorescein arsenical hairpin (FlAsH) reagent to label a natural substrate containing MA and the N-terminal domain of CA, Swanstrom's group [23] studied the effect of substrate context on processing the HIV-1 PR Gag and Gag-Pol polyproteins. These authors [23] observed that complex substrate interactions both beyond the active site of the enzyme and across the scissile bond contribute to defining the rate of processing by the HIV-1 PR. The rate of cleavage by this viral protease can also be inhibited or enhanced through the replacement of the P1 residue of the Gag processing sites [61].

2.3. Natural Substrate Recognition by Drug-Resistant HIV-1 PR Variants

One of the major challenges in enzyme-based drug design is the development of drug resistance, of which the clinically approved antiretroviral drugs are not exempted. HIV-1 PR drug resistance currently involves about 46 mutations, [64] and 69 HIV-1 PR mutant crystal structures are found in the RCSB PDB [18]

With the observed high level of drug resistance/multidrug resistance caused by HIV-1 PR mutants, this overview will be incomplete without addressing how mutated protease recognizes the natural Gag and Gag-Pol polyproteins segments. The impact of mutated HIV-1 PR on the recognition of their natural substrate precursor (NC-p1 cleavage domain) was studied using GRASP software, [65] and it was observed that most active site mutations are rarely in contact with substrates, but are crucial to the binding of inhibitors [40]. The Gag and Gag-Pol polypeptide precursors cleavage sites are recognized by this protein cleaving enzyme despite point mutation or insertion (such as HIV-1 subtype C-I36T↑T containing 200 amino acid residues [12]) due to the ability of the HIV-1 PR to develop drug resistance to inhibitors at the active site, thereby recognizing the shape of its natural substrates. Hence, inhibitors designed for HIV-1 PR and their mutants may easily undergo drug resistance, whereas the protease recognizes its polypeptide natural substrates and proceeds with its cleavage bioactivity [40].

The output from this computational simulation [40] was followed by a crystallographic investigation conducted at 1.44–2.10 Å resolution [41]. The main aim of this study was to describe the substrate recognition mechanism by drug-resistant HIV-1 PR with a co-evolutional natural substrate (NC-p1) whose processing is not just the slowest, but also, the rate-limiting cleavage step in the maturation of Gag [41]. The NC-p1 cleavage domain has been noted to coevolve with drug-resistant V82A HIV-1 PR mutation whereby this substrate's P2 amino acid residue mutates from an alanine to a valine in response to this mutated variant (Fig. **3**) [40]. Amazingly, their [41] maiden experimental study revealed that one of the HIV-1 PR flaps displayed a closed conformation while the other flap displays a discrete intermediate conformation [41]. P3-P1 residues of the substrate were also observed to be of great importance in substrate recognition [40, 41].

Tie *et al.* [66] carried out a high-resolution crystallographic analysis to elucidate substrate recognition by wild-type HIV-1 PR, V82A and I84V variants at the molecular level. CA-p2, p2-NC, NC-p1, p1-p6 and $p6^{pol}$-PR cleavage domains were co-crystallized with these two drug-resistant mutants and wild-type PR to examine their conformation changes and estimate kinetics. The substrate sequence was six to eleven amino acids long. The 3-D X-ray crystal structures, which were refined to 1.1–1.6 Å resolution, revealed that the binding affinity and recognition features of these mutant PRs are partly dependent on the substrates' conformational flexibility. The natural substrates have more adaptability to bypass new conformations induced by drug-resistant mutants than potential inhibitors.

In order to confirm this coevolved substrate with mutant HIV-1 PR and to examine in detail their substrate recognition pattern, Schiffer and coworkers [67] carried out another investigation at the molecular level.

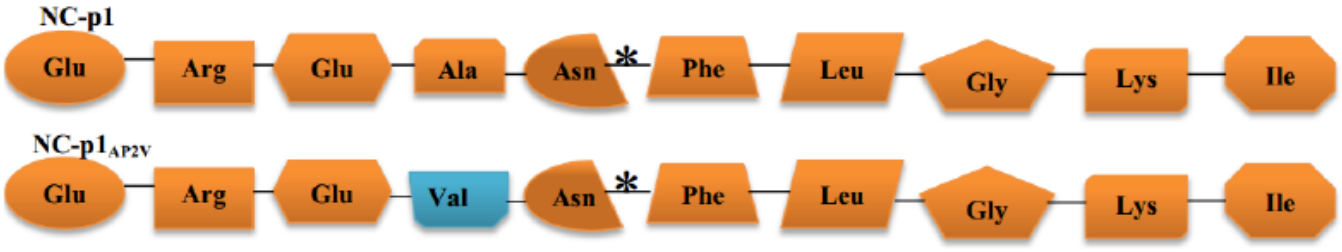


Fig. (3). Point mutation at P2 amino acid in NC-p1 natural Gag polyprotein cleavage domain. The scissile bond is indicated with asterisk and mutated amino acid is in green.

Three coevolved HIV-1 PR—natural substrates were studied; ^{AP2V}NC-p1$_{V82A}$, $^{LP1'F}$p1-p6$_{D30N/N88D}$ and $^{SP3'N}$p1-p6$_{D30N/N88D}$ in which two of these complexes have double mutations induced by nelfinavir treatment [68, 69]. The molecular modeling and dynamics analysis revealed that the conserved envelope conformation is an important feature for the HIV-1 PR mutants to recognize and cleave their substrates [67]. Closely related research was again reported by the same group [70] using an Anisotropic Network Model (ANM) [71]. In addition to the substrate envelope theory with respect to drug-resistant HIV-1 PRs and coevolved substrates, these authors [70] proposed that the recognition of the asymmetric substrates by the HIV-1 PR and its known mutants may likely be a result of the flexibility in the flap regions, which control the intrinsic dynamics.

So far, all natural target sequences documented herein do not exceed P5 to P5′. Using both *in silico* and *in vitro* techniques, Laco [72] put forward a novel analysis on substrate recognition by native, drug-resistant and multidrug-resistant (MDR) HIV-1 PRs in an attempt to derive the significance of the substrate-groove (S-groove) to natural target recognition by these PRs [72]. The natural substrate (with 24 amino acid residues) was employed for this study in which MA-CA and NC-p1 cleavage domains were examined. From his numerous observations [72] a condensed summary is provided while focusing on substrate recognition of MDR3761 HIV-1 PR with seven protease inhibitor (PI) resistance mutations and four polymorphic changes with respect to the wild-type (WT) PR: [73]. (1) The S-groove can be described as an active pocket within the HIV-1 PR contributed by the two dimers, which allow the PR to bind up to 24-residues of a Gag and/or Gag-Pol cleavage sites. (2) The S-groove can increase HIV-1 PR affinity for substrate recognition and cleavage process. (3) MDR PR is highly dependent on S-groove contacts for substrate recognition, binding and cleavage. (4) When Gag polyprotein was used as the substrate, a similar bioactivity was observed for both native and MDR PRs [72]. In his recent investigation, Laco [72, 74] concluded that PRs interact with the Gag and Gag-Pol cleavage sites *via* the active-site and S-grooves [74].

Lately, the nine cleavage domains of the natural Gag and Gag-Pol precursors were co-crystallized with MDR769 HIV-1 PR by Kovari and co-workers [73, 75-77]. In one of their studies, a wide-open flap and an extended substrate pocket were observed in MDR769 HIV-1 PR [75]. This MDR PR has ten of its residues mutated and each natural target has seven amino units from P3-P4′ [75]. It was argued that the enzyme recognizes its substrates through their preserved shape facilitated by conserved water molecules and hydrogen bonds [75, 77]. It was noted that the essential water molecule between the substrate and flap tips (Fig. **2**) was absent in the reactive region of MDR769 HIV-1 PR [75] These prompted the authors [77] to postulate the probability of variations in the enzymatic mechanism of these mutant with respect to the WT PR.

Deshmukh *et al.* [78] have associated the ability of the protease to naturally cleave known regions within the Gag polyprotein sequence with the conformational dynamics of the protease flaps that cover the active site. This was elucidated using chemical exchange-based NMR spectroscopy approach and mutant proteases. They [78] upheld that "the protease flaps are actively involved in substrate recognition and regulate the lifetime of productive complexes, allowing protease to differentiate between scissile and non-scissile sequences".

Understanding the recognition profile at the molecular level for the natural targets of HIV-1 PR and resistant mutants is crucial to the development of more active inhibitors. Many researchers have used computational techniques to provide a plausible description of this event in which more advanced *in silico* approach might likely yield better hypotheses.

2.4. Thermochemistry and Kinetic Parameters for HIV-1 PR and its Natural Substrates

Many *in vitro* analyses have been carried out to establish the kinetics parameters for HIV-1 PR and its natural substrate. These parameters are important factors in determining the catalytic activity of an enzyme and its substrate/inhibitor. Earlier kinetic constants found in the literature for HIV-1 PR and its natural targets were documented [30]. We therefore, present in Table **2**, a summary of inhibition constants and binding free energies from literature. Kinetic constants for mutated variants of HIV-1 PR are also included and all values presented are outputs involving sequences consistent with natural targets.

3. REACTION MECHANISMS OF HIV-1 PR AND ITS NATURAL SUBSTRATES

In recent times, studies on reaction mechanisms have gone far beyond what happens when a mixture of constituents yields one or two distinct product(s). Investigating the reaction mechanism is imperative to understanding enzyme catalysis, design of drugs and

Table 2. Experimental kinetic parameters for the proteolysis of natural Gag and Gag-Pol cleavage domains by HIV-1 protease and its mutants.

Cleavage Site	Peptide Sequence	HIV-1 PR	ΔG_{bind} kcal mol^{-1}	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($mM^{-1}.s^{-1}$)	K_d
MA-CA	SQNY*PIVQ	WT		3.75 [79] 0.78 [80]	23.00 [79] 15.70 [80]	6.20 [79] 20.10 [80]	
MA-CA	VSQNY*PIVQ	WT		0.15 [81] [48, 80]	6.90 [81] 6.80 [48]	46.00 [81] 45.30 [82]	
MA-CA	QNY*PIVQ	WT	-8.33 [75]	0.91 [80]	1.60 [80]	1.81 [80]	8.20e-7 [75]
MA-CA	VSQNY*PIV	WT		0.12 [80]	7.90 [80]	65.80 [80]	
MA-CA	SQNY*PIV	WT		0.53 [80]	13.50 [80]	25.40 [80]	
MA-CA	QNY*PIV	WT		0.86 [80]	1.00 [80]	1.20 [80]	
MA-CA	VSQNY*PIVQ	M46L		0.56 [81]	18.40 [81]	32.90 [81]	
MA-CA	VSQNY*PIVQ	V82S		1.34 [81]	13.40 [81]	10.00 [81]	
MA-CA	VSQNY*PIVQ	V82A		0.42 [81]	7.00 [81]	16.70 [81]	
MA-CA	VSQNY*PIVQ	I84V		1.02 [81]	93.50 [81]	92.00 [81]	
MA-CA	VSQNY*PIVQ	L90M		0.64 [81]	33.00 [81]	51.60 [81]	
MA-CA	QNY*PIVQ	MDR769	-2.02 [75]				3.40e-2 [75]
CA-p2	ARVL*AEAM	WT		0.37 [79]	2.60 [79]	6.90 [79]	
CA-p2	RVL*AEAM	WT	-9.14 [75]				2.10e-7 [75]
CA-p2	RVL*AEAM	MDR769	-2.02 [75]				3.40e-2 [75]
p2-NC	ATIM*MQRG	WT		3.00 [79]	41.00 [79]	14.00 [79]	
p2-NC	TIM*MQRG	WT	-9.14 [75]				2.10e-7 [75]
p2-NC	TIM*MQRG	MDR769	-5.24 [75]				1.50e-4 [75]
NC-p1	ERQAN*FLGKI	WT		0.17 [81]	0.15 [81]	0.90 [81]	
NC-p1	ERQAN*FLGKI	V82A				0.70 [81]	
NC-p1	ERQAN*FLGKI	I84V					
NC-p1	ERQAN*FLGKI	M46L				0.30 [81]	
NC-p1	ERQAN*FLGKI	V82S				0.40 [81]	
NC-p1	QAN*FLGK	WT	-6.74 [75]				1.20e-5 [75]
NC-p1	QAN*FLGK	MDR769	-3.83 [75]				1.60e-3 [75]
p1-p6	PGNF*LQSR	WT		>5.00 [79]	<0.25 [79]	0.05 [79]	
p1-p6	RPGNF*LQSRP	WT		1.20 [81]	0.98 [81]	0.80 [81]	
p1-p6	RPGNF*LQSRP	V82A				1.20 [81]	
p1-p6	RPGNF*LQSRP	V82S				1.20 [81]	
p1-p6	RPGNF*LQSRP	I84V				1.20 [81]	
p1-p6	RPGNF*LQSRP	M46L				0.60 [81]	
p1-p6	RPGNF*LQSRP	L90M				2.10 [81]	
p1-p6	GNF*LQSR	WT	-5.33 [75]				1.30e-4 [75]

(Table 2) contd….

Cleavage Site	Peptide Sequence	HIV-1 PR	ΔG_{bind} kcal mol^{-1}	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($mM^{-1}.s^{-1}$)	K_d
p1-p6	GNF*LQSR	MDR769	-3.52 [75]				2.70e-3 [75]
TF-PR	FQF*PNIT	WT	-8.84 [75]				3.50e-7 [75]
TF-PR	FQF*PNIT	MDR769	-3.02 [75]				6.30e-3 [75]
PR-RT	LQF*PISP	MDR769	-3.42 [75]				3.20e-3 [75]
RT-RH	AETF*YVDG	WT	-6.91 [83]	0.19 [79]	1.80 [79]	10.00 [79]	6.21e-6 [83]
RT-RH	ETF*YVDG	WT	-6.44 [75]				2.09e-5 [75]
RT-RH	ETF*YVDG	MDR769	-2.23 [75]				2.40e-2 [75]
RH-IN	RKIL*FLDG	WT		1.15 [79]	5.10 [79]	4.40 [79]	
RH-IN	KIL*FLDG	WT	-9.68 [75]				6.40e-8 [75]
RH-IN	KIL*FLDG	MDR769	-6.53 [75]				1.70e-5 [75]

Note: PR is the protease, ΔGbind is the intrinsic binding free energy, Km is the Michaelis constant, Kcat is the catalytic constant, Kcat/Km is the rate of catalysis, Kd and Ki denote the dissociation and inhibition constants, respectively.

their metabolism. Here, we address the reaction mechanism of the native HIV-1 PR with natural Gag and Gag-Pol polyprotein cleavage domains. Contributions from the use of both experimental and theoretical techniques are presented on the mechanistic aspect of this enzyme with its substrates. Several key factors that contribute to the reaction mechanisms of HIV-1 PR and its natural substrates, such as, the protonation state of the aspartate dyad, the nature of the reaction, reaction process and conditions, role of water molecule(s) and finally, energetics are discussed.

3.1. Effect of Protonation State on the Catalytic Aspartate Dyad

Placement of proton(s) on the Asp dyad of HIV-1 PR is crucial to understanding the reaction mechanism and is still a subject of debate [10, 84]. A review describing the protonation state of the ionizable group of aspartate proteases has appeared recently [34]. Experimental techniques such as combined X-ray/neutron crystallography, [85, 86] NMR spectroscopy, [87-89], calorimetric and binding kinetics [34] were used to determine the protonation state in aspartic proteases. Similarly, theoretical methods have also been applied to understand the ionization state of these catalytic Asps. These methods include; QM/MM, [26, 90-92] MD simulations [93-95] and online software (such as PROPKA, [96] PDB2PQR [97] and H++ [98]). One of the latest theoretical reports on the detailed protonation pattern of the Asp dyad is from Roitberg and coworkers, [99] where pH Replica-exchange MD (REMD) [95] was used to elucidate the pK_a of both catalytic and titratable groups of apo and inhibitor-bound HIV-1 PR. It was observed in the substrate-free model that the protonation of Asp25 and Asp25′ are chemically indistinct while the bound model exhibited monoprotonation. It was remarked that the binding of an inhibitor rarely induces change in the ionizable states of the catalytic dyad. However, this former moiety can invoke shifts in the pK_a values of the Asp dyad as observed from neutral asymmetric inhibitors that were studied [99]. A correct understanding of the protonation state(s) of the HIV-1 PR—substrate complex catalytic aspartates is crucial for accurate QM modelling of the reaction mechanism.

3.1.1. Possible C. of HIV-1 PR for Catalytic Function

Evidence from both experimental and theoretical analyses has revealed that the catalytic process of HIV-1 PR functions at pH of 2–7 [33, 34, 100]. This process usually occurs at the active site, which consists of two coplanar aspartate moieties and is usually stabilized by the catalytic triad Asp25(25′), Thr26(26′) and Gly27(27′) peptide units [101]. The Asp 25(25′) dyad of the free enzyme can exist in monoprotonated [99] (four possibilities, model **A-D**), an unprotonated [102, 103] (model **E**) and diprotonated [104, 105] (four possibilities, model **F-I**) forms (Fig. **4**).

Piana and Carloni [100] reported the observation of a low barrier hydrogen bond (LBHB) within the apo HIV-1 PR of reactive Asp monomers (model **J**, Fig. **4**). These researchers [100] utilised *ab initio* simulations at 300 K to highlight the requirements for this enzyme to carry out its catalytic functions at the active site. These include, monoprotonation specifically on OD1 of Asp25′, a catalytic water (WatC) molecule close enough to the carboxyl oxygen of the Asp moieties and the contribution of Thr26(26′)–Gly27(27′) to the sta-

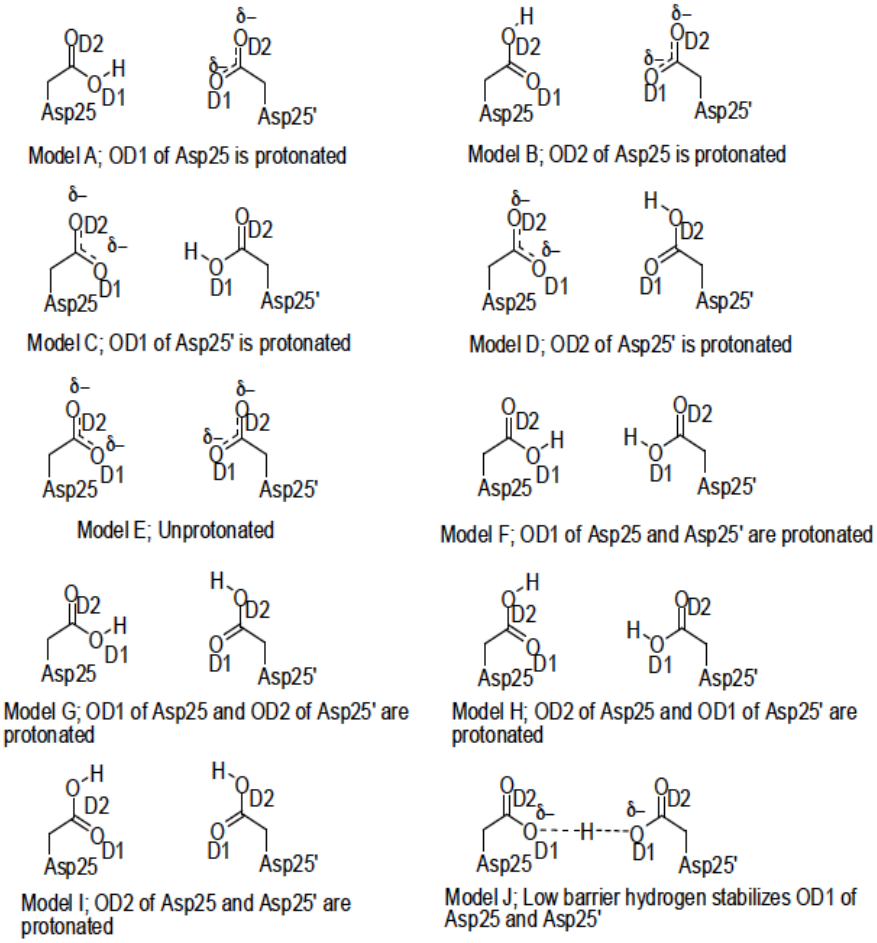


Fig. (4). Proposed protonation models for the HIV-1 PR catalytic dyad [100, 105, 106, 109].

bility of the Asp dyad orientation [100]. Furthermore, the catalytic reactivity of HIV-1 PR—substrate complex is dependent on the conformational flexibility of the enzyme [25, 106, 107]. The dynamics of the flap region of HIV-1 PR allows the inflow of the substrate and the release of the cleaved products. When investigating these interactions with *in silico* methods, an accurate theoretical level and large basis set are recommended to obtain reasonable geometry orientation and energetics [1, 25].

The catalytic core domain of the PR must attain a specific geometry before it can process its substrate/inhibitor's functionality, Okimoto *et al.* [93] termed this specific conformation the "active conformation". The proximity of the enzyme with the substrate in the presence of WatC and protonated OD2 of the Asp25′ residue to form the enzyme-substrate (ES) complex should be within specific distances (Fig. **5**). Using the Hartree-Fock (HF) [108] theoretical method and the 6-31G(2d) [109] basis set, these authors [93] determined the necessary conditions for an active conformation to enable HIV-1 PR to cleave its natural substrates. These conditions are: (a) the two catalytic Asp dyad should hold WatC through formation of hydrogen bonds (Fig. **5**: d1, d2, d3, and d4), (b) the interatomic distance between the oxygen of WatC and the scissile carbonyl carbon of the substrate should be maintained within 3.3 Å (Fig. **5**: d5) and this was based on the VDW radii of the carbon and the oxygen in retrospect, [93] (c) and the oxygen of the cleavable carbonyl forms a hydrogen bond with the OD2 of Asp25′ (Fig. **5**: d6).

The effect of conserved water molecules on the reactive conformational state of the enzyme was also addressed therein [22, 110]. The 3-D arrangement of dipoles from these structurally conserved water molecules contribute to the favorable geometric of this enzyme required for its cleavage bioactivity [110]

3.2. The Nature of the Reaction

The mechanistic reaction of the HIV-1 PR has been extensively studied using both theoretical and experimental techniques and categorized either as a nucleophilic [1, 111, 112] or a general acid-base [1, 111, 113-

120] reaction. The major distinction between these two mechanisms is the presence of a water molecule in the latter and absence in the former. Although the general acid-base mechanism is widely accepted, the nucleophilic reaction cannot be entirely ruled out. Within the context of this review, we highlight these two reaction mechanisms based on contributions from *in silico* and *in vitro* perspectives.

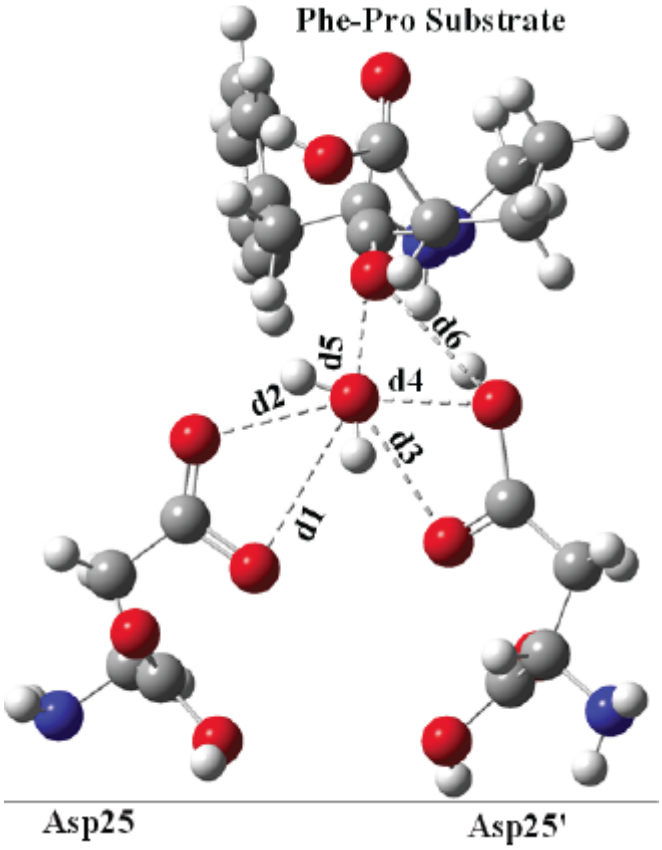


Fig. (5). Interatomic distances obtained with *ab initio* HF/6-31G(2d) for the ES complex, redrawn from literature.[94] Carbon, oxygen and nitrogen atoms are represented with grey, red and blue, respectively. All distances are in Å, d1=3.28, d2=2.66, d3=2.96, d4=2.80, d5=2.91 and d6=2.97.

3.2.1. Nucleophilic Reaction

Earlier thoughts about the hydrolysis of natural substrates by HIV-1 PR were that it most likely proceeds through a direct nucleophilic attack by an aspartate [111]. This direct nucleophilic attack process involves the catalytic aspartate group whereby the unprotonated Asp acts as the nucleophile rather than a water molecule. The nucleophile generally attacks the carbonyl carbon of the scissile amide. This reaction has been studied to occur as a stepwise or a concerted nucleophilic reaction mechanism.

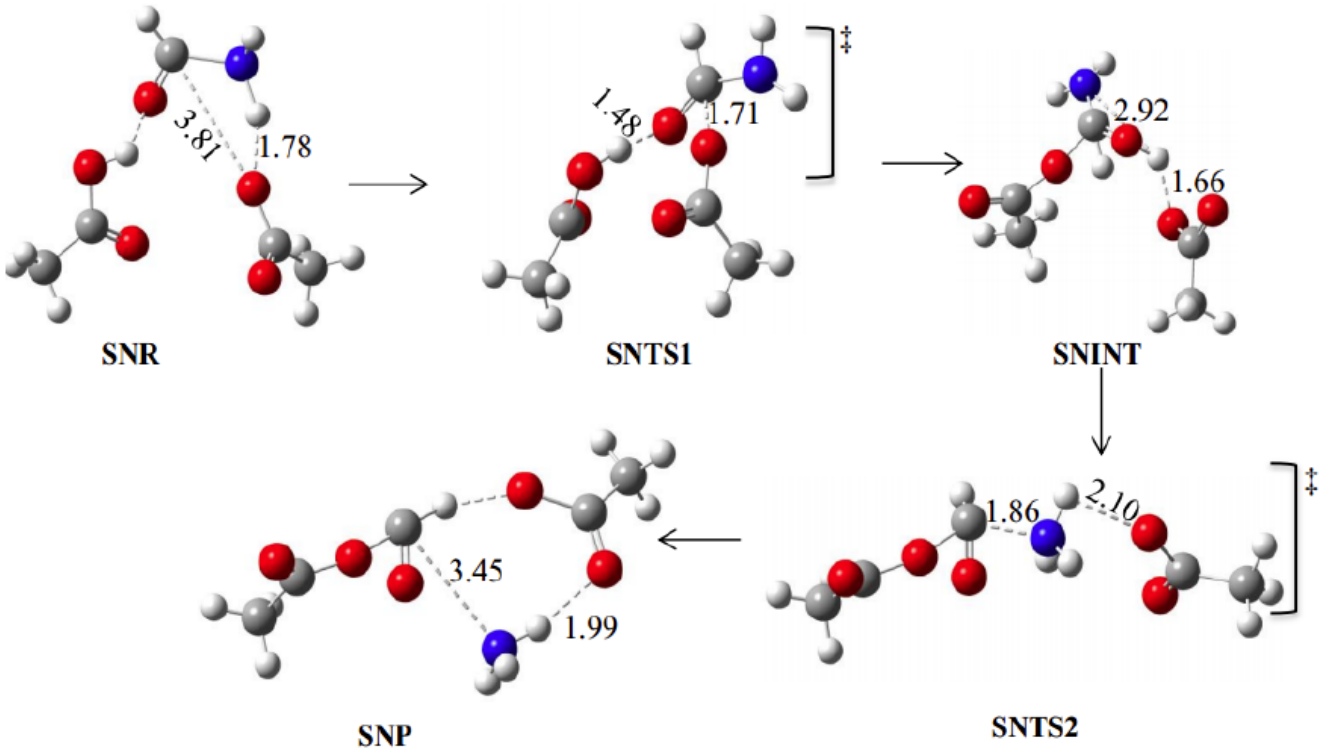
3.2.1.1. The Stepwise Nucleophilic Mechanism

The nucleophilic mechanism for HIV-1 PR catalysis was studied to be stepwise whereby the reaction proceeds through more than one TS structure and intermediates. The nucleophilic attack of the unprotonated Asp at the scissile carbonyl carbon produces an anionic intermediate that abstracts a proton from the protonated Asp (Scheme **1**). The resulting acyl-enzyme intermediate is then hydrolyzed, with the unprotonated Asp acting as a general base [111]. The stepwise nucleophilic mechanistic pathway was explored herein [1, 111, 112] using computational methods. Both quantum mechanics (QM) and MD techniques have been harnessed to give a total account of the reaction process involved in this mechanism. In addition to their exploration of all possible mechanistic pathways for the enzymatic mechanism of HIV-1 PR, Park *et al.* [1] investigated the stepwise nucleophilic mechanism (Scheme **1**) at MP2/6-31G(2d)//RHF/6-31G(2d) level of theory. The reaction steps involved the formation of a reactant complex, two TS structures, one intermediate and product complex. An overall $\Delta G^{\ddagger}$ of 36.6 kcal mol^{-1} was estimated for this mechanism.

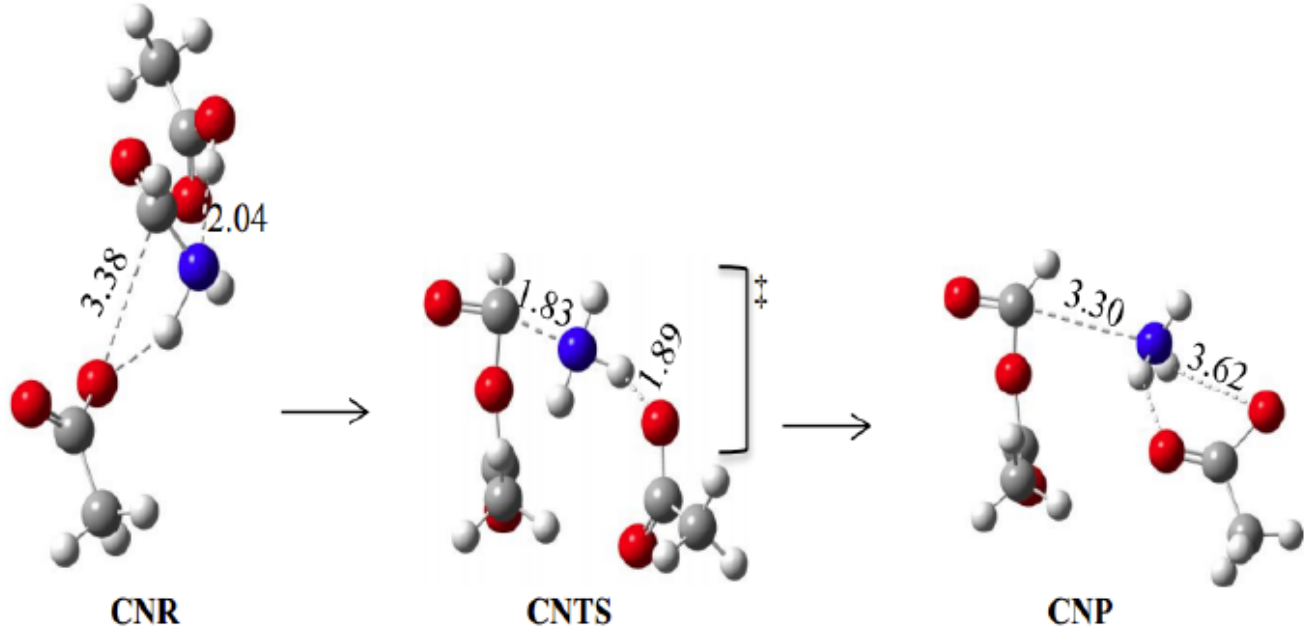
Approximate valence bond (AVB) [121] method was parameterized and tested on the cleavage of the natural substrate by the HIV-1 PR [122]. The AVB formalism was developed based on density functional theory (DFT) [123] approach and it allows partitioning of the exact atoms involved in the reaction and other surrounding amino acid residues and solvents. Atoms at the reaction center [122] were modeled using AVB while the rest of the system was treated with a classical MD [54] method. This computational method allows the authors [122] to investigate the stepwise nucleophilic mechanism in which protonation was assigned to the two Asp group (model **I**, Fig. **4**). The first stage of the mechanism involves proton transfer between the inner oxygens (OD1) of both Asp and subsequent nucleophilic attack by Asp25. As a result of the relatively short simulation time for AVB/MM technique employed by these authors, [122] it was impossible to estimate the energy barriers of the entire nucleophilic process.

3.2.1.2. The Concerted Nucleophilic Mechanism

The mechanistic pathway for HIV-1 PR catalysis of substrates could also be concerted nucleophilic, which proceeds with the formation of an enzyme-substrate complex leading to a concerted nucleophilic TS structure (Scheme **2**). According to Park *et al.* [1] investigation, the concerted nucleophilic reaction involves the concurrent attack of the nucleophile (unprotonated

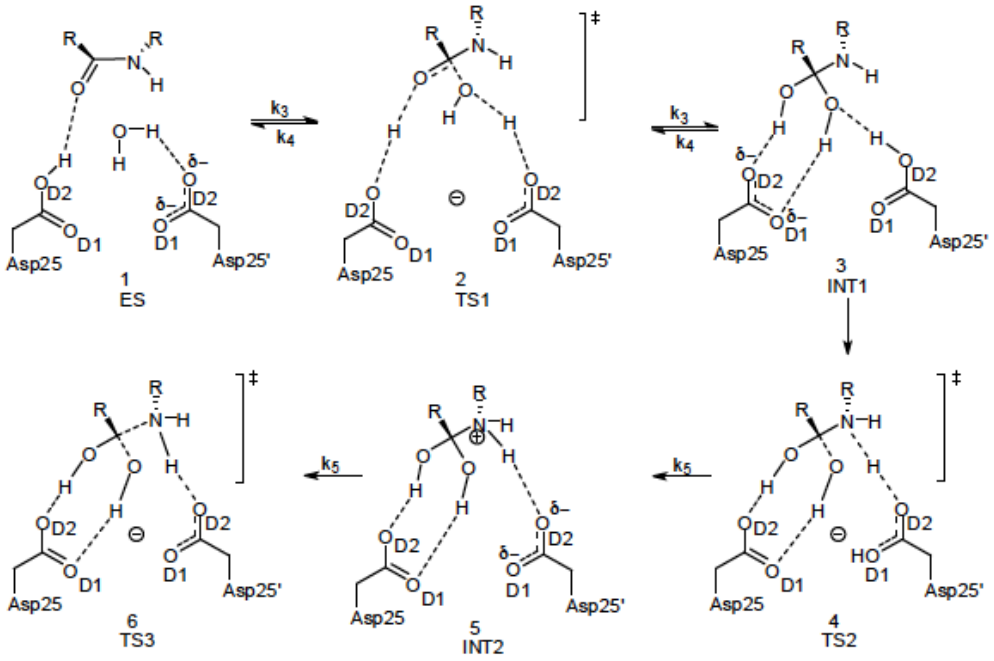


Scheme 1. A pictorial representation of the structures involved in stepwise nucleophilic mechanism for the hydrolysis of a simplified peptide model by HIV-1 PR using MP2/6-31G(2d)//RHF/6-31G(2d).[1] The legends are: SNR = stepwise nucleophilic reactant, SNTS1 = stepwise nucleophilic transition state 1, SNINT = stepwise nucleophilic intermediate, SNTS2 = stepwise nucleophilic transition state 2, SNP = stepwise nucleophilic product. All distances are in Å.



Scheme 2. Representation of the structures involved in concerted nucleophilic pathway for the hydrolysis of a simplified peptide model by HIV-1 PR using MP2/6-31G(2d)//RHF/6-31G(2d).[1] The legends are: CNR = concerted nucleophilic reactant, CNTS = concerted nucleophilic transition state, CNP = concerted nucleophilic product. All distances are in Å.

Asp) on the carbonyl carbon and transfer of a proton from the acidic Asp to the scissile nitrogen of the substrate. The nucleophilic Asp bound tightly to the dissociated scissile carbonyl in the concerted nucleophilic product (Scheme **2**). In their study, Park *et al.* [1] used *ab initio* calculations including Møller-Plesset second-



Scheme 3: Proposed mechanism for the substrate hydrolysis reaction catalysed by HIV-1 protease initiated by Asp25 protonation. Structures along the pathway are indicated as follows: (1) enzyme—substrate complex; (2) water attack TS; (3) tetrahedral gem-diol intermediate; (4) scissile N-protonation TS; (5) protonated amide intermediate; and (6) cleavage of scissile C-N bond TS leading to separated product complex. Redrawn from literature [26, 124, 132].

order perturbation (MP2) [124] and HF [108] theories coupled with the 6-31G(2d) [109] basis set, to describe the concerted nucleophilic mechanism. They used a model reaction where formamide represents the substrate, while acetate and acetic acid represent the catalytic Asp residues [1]. For this reaction, an activation free energy ($\Delta G^{\ddagger}$) of 25.8 kcal mol^{-1} was observed [1].

3.2.2. The General Acid-Base Mechanism

The general-acid base mechanistic pathway for HIV-1 PR proteolysis of its substrates came into limelight in the early 1990s [125, 126] following previous experimental research on related proteolytic enzymes [127, 128]. Recent evidence weighs strongly in favor of the general acid-base catalysis through the use of combined X-ray/neutron crystallography, [85, 86] to detect notable reaction complexes along the PES of the HIV-1 PR—natural target. This was earlier proposed to be rarely possible due to the reactive nature of the protease [1, 129]. In this mechanism, a water (WatC) molecule acts as the nucleophile within the active site of the HIV-1 PR. The catalytic process has been proposed to either be stepwise or concerted.

3.2.2.1 Stepwise General Acid-Base Mechanism

The stepwise general acid-base (SG) mechanism is the most widely studied model for the catalysis of HIV-1 PR. The reaction is quite diverse and puzzling opinions continued to appear as researchers try to unravel the mechanism using both *in silico* and *in vitro* methods. Important research questions are: Can a single protonation pattern be sufficient for the proteolysis of all the natural substrates? How many TS structures are involved in SG mechanism and what is the rate-determining step?

Within the framework of this review, we have considered studies together based on the protonation pattern (Fig. **4**) used/proposed by different contributors from both computational and experimental investigations. This becomes a better approach due to the importance of protonation on the catalytic Asp group to enable HIV-1 PR enzymatic functions. It is quite interesting to observe that most recent reports on HIV-1 PR—substrate catalysis has adopted a monoprotonation mechanism (models **A-D**; Fig. **4**).

3.2.2.1.1. The Stepwise General Acid-Base Mechanistic Pathway of HIV-1 PR Involving Asp25 Protonation

Protonation assignment to the catalytic Asp group has been studied to exist in a number of possible models (Fig. **4**). On the active Asp25, two oxygens of the carboxyl group could either be protonated for HIV-1 PR catalytic function. For ease of readability, the discussion is presented in the light of the applied *in silico*

and *in vitro* methods, the perspective of the study and major contributions towards describing the mechanism.

To give a total account of the reaction process involved in enzyme catalysis, the approximate valence bond (AVB) [121] method was parameterized and tested on the cleavage of the natural substrate by the HIV-1 PR [122]. Protonation of the catalytic groups was specifically on OD2 of Asp25 (model **B**, Fig. **4**). Shown in Scheme **3** is the adopted reaction model by Trylska *et al.* [122]. Each reaction step was elucidated using HIV-1 PR—p2-NC cleavage mechanism with the scissile bond at Met*Met (Table **1**). It was summarized that the AVB/MM technique would allow a detailed description of the enzymatic reaction. This reaction steps include; intermolecular proton transfers from water to aspartate, nucleophilic attack of the reduced water on the substrate scissile bond, conformational changes of the tetrahedral intermediate, two proton transfers between the intermediate and the Asp residues, and cleavage of the intermediate into reaction products [122]. As observed in the nucleophilic mechanism, the short simulation time hindered the possibility of estimating the energy barriers of the general acid-base mechanism using AVB/MM method [122]

Trylska *et al.* [112, 119, 122] continued their research by providing a broader understanding on the enzymatic reaction of HIV-1 PR through the application of AVB/MM MD simulations. Protonation pattern **B** (Fig. **4** and Scheme **3**) was found to be the most suitable model for the general acid-base enzymatic catalysis of HIV-1 PR on the *N*-acetyl-Thr-Ile-Met*Met-Gln-Arg-amide sequence [112]. The role of hydrogen bonding in the reaction mechanism of the enzyme and one substrate was also elucidated. The hydrogen bonds observed were characterized as; [112] short and strong, weak and long, centred or asymmetric and single-well hydrogen bonds. It was observed that strong hydrogen bonds are peculiar to both mechanistic pathways (nucleophilic and general acid-base) and assist in simultaneous proton transfers within the reactive field. A low barrier and central hydrogen bond were noted between the scissile NH and the OD2 of Asp25′. OD1 of both Asp dyad formed an asymmetric-LBHB that facilitates the stability of the coplanar aspartates interchange between them. For the cleavage to occur, the nitrogen atom of the scissile bond should be protonated. Though computationally expensive, the authors [112] suggested a technique that explores the full conformational space such as umbrella sampling [130, 131]. This method gives an overt free energy profile capturing all possible moieties along the reaction coordinate.

Over time, new improvements were put forward to establish a suitable mechanistic model for HIV-1 PR catalysis of its substrates. Taking into consideration the flexibility and electric field of the enzyme plus active site interaction with other residues as suggested herein, [106, 107] Rothlisberger and coworkers [25] offered a hybrid Car-Parrinello (quantum mechanics) and molecular mechanics (QM/MM) method [132] at 15 ps to describe the chemical steps involved in the cleavage of P3–P3′ sequence of p2-NC by HIV-1 PR. Using the BLYP hybrid exchange [133, 134] DFT formalism and the plane wave (PW) [135] basis set for the QM part, these authors [25] investigated two possible protonation patterns (models **B** and **J**, Fig. **4**) for the ionizable aspartates at the active site.

Through this all-atoms simulation, the nature of each moiety along the reaction profile was discussed extensively with respect to reaction energies. Two plausible models were also explored for the formation of intermediate (INT); the gem-diol and the oxyanion (Fig. **S1**; **SI**), which was proposed to be possible through a tunnelling [136] effect. The required distance for tunnelling for the stability of an oxyanion INT was not achieved, but it was suggested that increasing the simulation time may yield a positive result [25]. The $\Delta G^{\ddagger}$ for the formation of gem-diol INTs was 18 kcal mol^{-1} *via* model **B** protonation mechanism. The data suggest that the catalytic reaction is more favorable when OD2 of Asp25 is protonated. Even when the reverse reaction was examined the same free energy barrier was obtained, [25]. and including the catalytic triad in the QM part to monitor the reaction mechanism gave the activation free energy of 15 kcal mol^{-1} [25]. This lower energy difference in moving from ES to INT was rationalized with respect to the catalytic strength of HIV-1 PR which is partly driven by the polarity of the cleavage site surrounding [107]. When the alpha carbon (Cα) atoms of this enzyme were kept frozen, the calculated free energy for this same step was 25 kcal mol^{-1}, [25] an indication of the vitality of flexibility for reactive catalysis [25, 107]. The theoretically predicted TSs energy value was 18 and 21 kcal mol^{-1} in which the second TS is 5 kcal mol^{-1} higher than experimentally deduced value [137, 138]. These authors [25] also carried out QM calculations to rationalize this discrepancy.

Two of our hypothetical questions form part of Bjelic and Åqvist [110] aims when they analyzed the catalysis of Plm II, Cat D and HIV-1 PR on their respective natural substrates. The reaction mechanism was studied with two natural substrate sequences namely; RH-IN and CA-p2 from the Gag and Gag-Pol

cleavage domains, respectively. The applied *in silico* method was a triple hybrid one in which empirical valence bond (EVB) [139, 140] was coupled with free energy perturbation (FEP) [139, 141] and classical MD simulations. This integrated approach can be used for enzyme catalysis to dissect the reaction steps, elucidate the features of the structures along the reaction coordinate and predict free energy values that are closer to experimental data [110]. After a detailed literature survey, protonation pattern **B** (Fig. **4**) and the general acid-base stepwise mechanism (Scheme **3**) were adopted for their [110] calculations. Elucidating the nature of each structure along the reaction coordinate, these authors observed a flat and almost similar in height tetrahedral INT structure between TS1 and TS2. The theoretical free activation energies for the enzymatic process on RH-IN were 17.2, 16.4 and 16.6 kcal mol^{-1} for TS1, INT and TS2, respectively, [110] in which the TS1 value is in agreement with the experimental value of 17.6 kcal mol^{-1} [113]. Apart from the active site WatC, a total of five water molecules was observed to be crucial for stability and reactivity, [110] which is supported by an earlier crystallographic study of this enzyme [22]. Despite not retaining these water molecules in the enzyme preparation procedure, electrostatic interaction energy arising from them and the INT structures in both substrates amounts to -12 kcal mol^{-1} [110, 142-144]. A contrasting observation was put forward by these authors on the notion that the HIV-1 PR intrinsic conformational flexibility is substantial to the reduction of the free energy barrier in the catalytic process [25, 145].

Kipp *et al.* [146] gave an answer to one of the research questions on the similarity between the mechanistic route of drug-resistant and the native HIV-1 PRs catalysis of substrates. They [146] reported nearly identical TS structures in both complexes. Exploring enzymatic transition state analogues enables the development of better inhibitors, a powerful tool for drug discovery [147]. Kinetic Isotope Effects (KIEs) [126] and ONIOM methods were used to elucidate the chemical mechanism of native and I84V mutant HIV-1 PR—Acetyl-Ser-Gln-Asn-Phe*Pro-Val-Val-NH_2 systems. The reaction was proposed [146] to have occurred in six phases (Scheme **3**) and TS2 was noted as the rate-determining structure as earlier proposed in the literature [25, 113, 118]. This TS corresponds to the protonation process of the substrate scissile nitrogen atom.

The relevance and implication of computational techniques in studying reaction mechanism can not be underestimated. Theoretical chemists can easily model a system and calculate the energetics in detail as well as locate the important moieties along the reaction coordinate. Most recently is the use of an advanced computational technique such as enhanced sampling that explores the full conformational space of the reaction process [112]. Recently, an investigation was done by Krzemińska *et al.* [120] on the catalytic reaction of HIV-1 PR with a non-natural substrate. The study utilizes an intriguing model in which both electrostatic and dynamic contributions were accounted for. The applied methods are sophisticated with the observation of about 14 reaction complexes that were revealed from the umbrella sampling method, QM/MM/MD and QM/QM/MM approach (Fig. **S1**; **SI**). Although the modeled system is a 6-Alanine peptide sequence (not an exact natural substrate), some highlights from their [120] analysis are provided.

Two mechanistic pathways involving OD2 protonation of both Asp25 and Asp25′ were investigated. The mechanistic route that initiates through the protonation of the outer oxygen of Asp25 (model **B**; Fig. **4**) leading to the formation of the gem-diol tetrahedral INT is the most plausible pathway. The breakdown of this INT into product fragments is likely the rate-determining step as earlier proposed through experiment [113] with $\Delta G^{\ddagger}$ of 15 kcal mol^{-1} [120]. This theoretical $\Delta G^{\ddagger}$ is in agreement with experimental values within 15 and 17.9 kcal mol^{-1} [125, 148]. A concerted process was observed for the plausible rate-determining step, which is characterized by two proton transfers. The first involves the transfer of hydrogen atom from Asp25′ OD2 to the substrate scissile nitrogen and the second hydrogen moves from a diol group of the tetrahedral INT to Asp25 OD2 (Scheme **3**).

These authors [120] also measured the enzyme kinetic isotope effects (KIEs) for the TS structures involved in the mechanism and obtained values which are in reasonable agreement with experimental values. Contributions from dynamic effect (inherent through enzyme fluctuation with respect to the reacting moieties [149]) on the activation barrier seem negligible compared to the electrostatic effect, which tends to be crucial for electronic reorientation of the geometry of molecules at the active site [120].

3.2.2.1.2. The Stepwise General Acid-base Mechanistic Pathway of HIV-1 PR Involving Asp25′ Protonation

The catalytic ability of HIV-1 PR was traced to its conformational adaptability by Piana *et al.* [107]. They [107] employed an integrated approach with *ab initio* Car-Parrinello MD simulation [132] (10.0 ps) and classical MD simulations (2.1 and 8.0 ns) at 300 K to in-

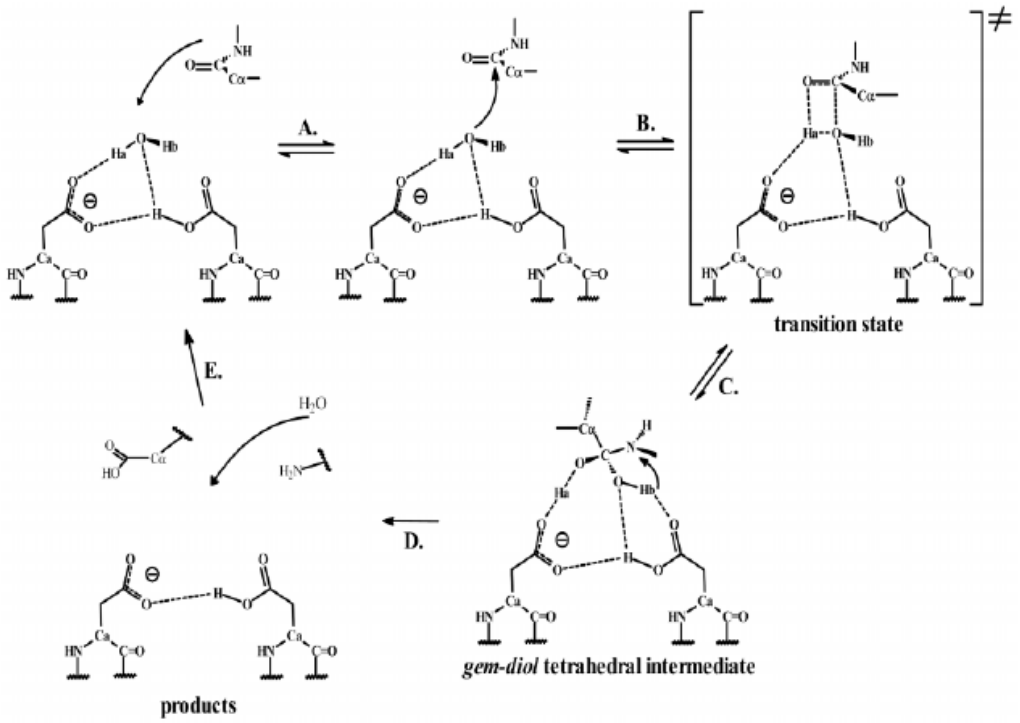
vestigate the HIV-1 PR—N-acetyl-Thr-Ile-Met*Met-Gln-Arg complex in which the substrate segments corresponds to six amino acid residues of p2-NC (Table **1**) [107]. The computational method was used to determine the cleavage sites polarization, conformational fluctuations and the reaction mechanism of the complex. The HIV-1 PR mechanistic interaction with its substrate was initiated as in Scheme **3**, with more emphasis on the first TS modeling leading to the tetrahedral intermediate formation [107] The authors [107] observed a mechanistic pathway which showcases OD1 of Asp25′ protonation (model **C**; Fig. **4**) as the energetically favorable compared to the rest of the models (**A**, **B** and **D**; Fig. **4**). Acyclic TS (similar to TS1 Scheme **3**) was observed from the *ab initio* calculations as a concerted process where the catalytic Asp dyad transfers a proton to the carbonyl substrate and accepts a proton from WatC. The $\Delta G^{\ddagger}$ of 21.5 kcal mol^{-1} obtained [107] was 5.5 kcal mol^{-1} higher than experimentally estimated values [125, 126, 150]. These authors [107] however, attributed this large difference in the theoretically predicted energetics to their modeled complex (propionic acid/propionate as Asp dyad, substrate as N-methyl acetamide while Thr26/26′ and Gly27/27′ N-methyl formamide) and possibly the applied theoretical approximation [134, 151]. It was concluded that the catalytic triad of the cleavage sites mainly induced the polarization of the reactants. The classical MD calculations revealed a large-scale protein fluctuation inherent from flaps and the cantilever, which modulates the conformational diversities of the substrates at the active site. They have also concluded that substrate residues further away from the active site may likely be vital to the substrate modulation [107].

Weber and coworkers [152, 153] proposed a perspective on the hydrolysis of substrates by HIV-1 PR and mutants I54V, V32I and I47V. The investigations involved both theoretical and experimental techniques. An autoproteolytic peptide sequence cleavage mechanism by WT, I54V, V32I and I47V HIV-1 PR mutants was investigated computationally after trapping notable moieties (through X-ray crystallography) along the reaction pathway of these PRs—peptide systems [152, 153] In both studies, the proposed mechanism of peptide bond cleavage was described as represented in Scheme **4**. The applied theoretical method, DFT; B3LYP/6-311++G(d,p), was used to provide more insight into the nature of the amino-gem-diol INT [152]. Protonation on the Asp dyad was observed to be favorable on OD1 of Asp25′ (model **C**, Fig. **4**). The PRs neither forms covalent bonds nor transfer its carboxyl hydrogen to the substrates [152] in contrast to Schemes **1-3** as revealed through experiment, [152] which will be discussed subsequently.

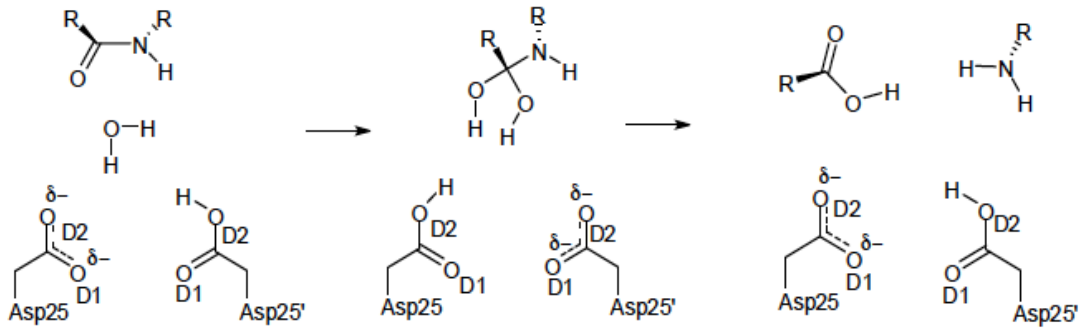
The proposed mechanism proceed with the Asp group participating as anchoring moieties (Scheme **4**) [152]. The reaction process involves formation of an enzyme—substrate complex (step A), nucleophilic attack of WatC on the scissile carbonyl with formation of 4-membered ring TS (step B), formation of gem-diol tetrahedral INT (step C), decomposition of the INT into product complex/protonation of scissile nitrogen (step D) and separation of the cleaved substrate from the enzyme (step E). The decomposition of the INT, which features protonation of the scissile nitrogen, was suggested to be the rate-determining step [152] as proposed from previous theoretical research [118].

Okimoto *et al.* [93] investigated the mechanism of this enzyme with the natural substrate using MD [54] simulations. They elucidated the catalytic and stabilizing role of water molecules around the loop structures of the HIV-1 PR. Two possible general acid-base hydrolytic mechanisms were proposed, Schemes **3** and **5**. The catalytic Asp dyad was monoprotonated in both mechanisms in which OD2 of Asp25′ (model **D**; Fig. **4** and Scheme **5**) was upheld to be ideal for the biological function of the PR. Simulating HIV-1 PR—PR-RT and HIV-1 PR—MA-CA complexes at 250 ps and 300 K in the presence of five characteristic water molecules, it was summarized [93] that the catalytic mechanism of HIV-1 PR initiates through the formation of enzyme—substrate complex involving a protonated outer oxygen of the Asp25′. The authors [93] suggested that the protonation state and water molecule(s) are crucial for the cleavage and processing of substrates to enable the replication of this virus.

Next, the same group [93, 94] monitored the effect of specific substrate sequences on the reaction mechanism of HIV-1 PR—substrate to identify efficient catalysis with respect to substrate nature [94]. This was studied using MD simulations in two different approaches; the no cut-off and the Particle Mesh Ewald (PME) [154] methods in which analyses were drawn at 300 K for 100 ps and 300 K for 250 ps, [94] respectively. OD2 of the Asp25′ residue was protonated and MA-CA cleavage domain was studied as six, seven and nine asymmetric sequences (Table **1**). An additional sequence in which Leu was substituted for Ser at P4 position within the nonapeptide was also studied [94]. In all cases, the reaction was proposed to have passed through a stepwise mechanism in which the reactants come together to form an ES complex and later pro-



Scheme 4. Proposed reaction mechanism of the peptide bond hydrolysis by HIV-1 PR devoid of the Asps bond sharing or atom exchange. "Adapted with permission from literature [153]. Copyright (2007) American Chemical Society."



Scheme 5. Summary of the mechanistic pathway involving Asp25′ OD2 protonation for the hydrolytic reaction of HIV-1 protease using an *ab initio* molecular orbital method.[94].

ceed to intermediate formation, substrate protonation and cleavage stages (Scheme **5**) [93, 94]. However, the last substrate sequence with a point mutation at the P4 position did not undergo catalysis due to its inability to fit in the active site [94]

Amongst the nine natural substrate cleavage domains, a scissile amidic bond is present at the Phe*Pro link in TF-PR and PR-RT cleavage junctions (Table **1**). Altoè *et al.* [26] described the catalytic process of HIV-1 PR—Phe*Pro complex to test the implementation of a novel QM/MM algorithm, tagged "Computations at Bologna Relating *Ab-initio* and Molecular Mechanics Methods" (COBRAMM). This approach is a hybrid model that acts as an interface within different programs through modularity. Through the combinations of programs, a user can easily design the desired computational level based on the aim(s) of the investigation [26]. Two distinct pathways were observed for this mechanism using QM and QM/MM approach. In the first method, only the reactive atoms within the enzyme's active site region and some neighbouring atom fragments were considered and apportioned into three layers. The applied DFT method involved B3LYP [108, 134] with DZVP, [155] 3-21G [156, 157] and STO-3G [158, 159] basis sets in the order high, medium and low layers, respectively. In the second (QM/MM) method, the partitioning is also similar (high and medium) while the last layer contains atoms treated with STO-3G (third layer in QM) and the rest of the HIV-1 PR residues (Fig. **6**) [26] which were assigned to AMBER force field (*ff03*) [160]. The protonated ionisable group of the reactive aspartates was

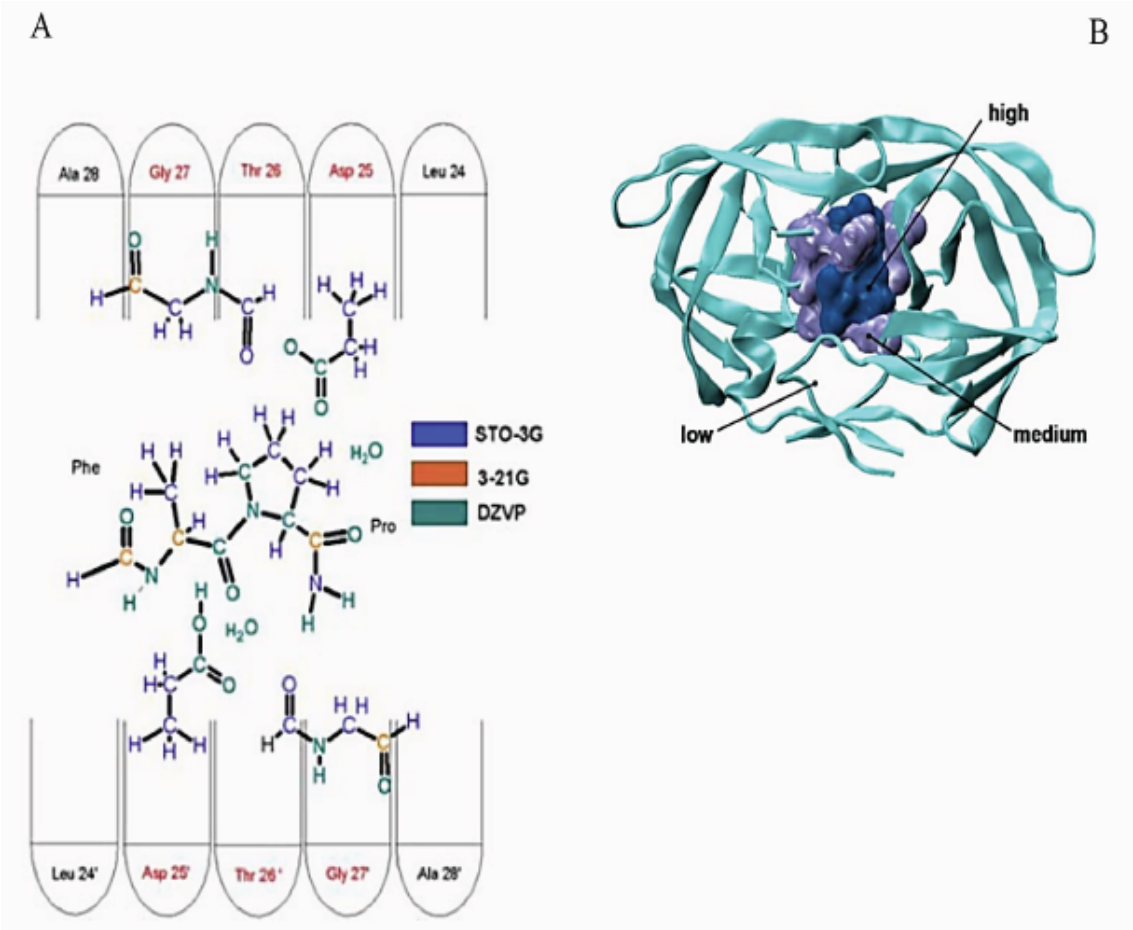


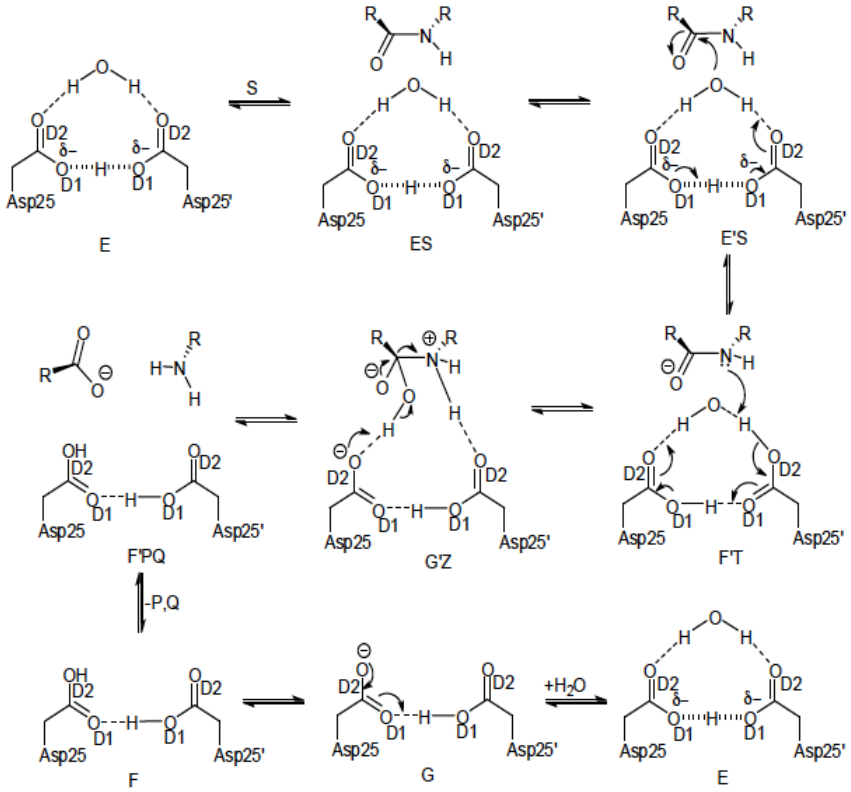
Fig. (6). A; The basis set adopted for the atoms of the QM system are shown using different colors. **B**: Model system used to study the QM/MM reaction profile of the enzyme HIV-1 PR. Pictures are adapted from literature [26] with permission "Copyright (2007) Springer-Verlag".

noted to be Asp25′ OD2 (model **D**; Fig. **4** and Scheme **5**).

From the QM calculation, four minima structures and two TS structures were observed along the PES for substrate hydrolysis [26]. One of the studied TS structures (corresponding to TS2 in Scheme **3**) corresponds to an earlier experimentally [113] and theoretically [118] proposed rate-limiting step with a characteristic highest peak along the reaction coordinate [26]. The activation free energy of approximately 23 kcal mol^{-1} is comparable to experimentally deduced results [125, 126]. The QM/MM model featured four different reaction moieties along its path with just one TS structure. The free energy barrier of this TS was predicted to be 21.4 kcal mol^{-1}, which is slightly lower than the QM data. However, the total process leading to product release was exergonic in QM and endergonic in QM/MM. Rationalizing these discrepancies as detailed as possible, the authors [26] concluded that the contribution of surrounding residues is crucial to product release. The proteolysis process was found to be more favourable in the QM method due to the smaller model system that was used (only the hydrolytic part was considered), while the QM/MM model system appeared to have restricted the system, *i.e.,* the flexibility of part/entire protein was hampered [26]

3.2.2.1.3. The LBHB Stepwise General Acid-Base Mechanistic Pathway of HIV-1 PR

Northrop [136] described a general chemical reaction process for the isomechanism of Asp proteases experimentally *via* the pH method. The mechanism was based on LBHB (model **J**, Fig. **4**) in which seven proton transfers were proposed for the catalytic turnover using a chromophoric substrate (Scheme **6**). Highlighting other related observations [161, 162] to provide exclusive understanding of the purpose of the LBHB in Asp proteases, Northrop [136] hypothesized that LBHB facilitates the formation of a cyclic complex. Formation of cyclic moieties along the reaction coordi-



Scheme 6. Chemical and kinetic isomechanism of an aspartic protease redrawn from literature.[109] E=Asp dyad and WatC at the enzyme's active site held with LBHB, ES=enzyme—substrate complex, E′S and F′T=interaction within the ES complex, G′Z=the initiation of the cleavage process of the substrate, F′PQ=the cleaved peptide, F and G are resonance structures of the Asps leading to the formation E to initiate another cycle.

nates in a given chemical process has been a subject of theoretical research in our group for over a decade [91, 163-169]. In aspartate proteases, cyclic complexation allows the distribution of electron density which stimulates the rate acceleration through reactant-state tunneling [136]

In addition to studying other stepwise general acid-base mechanism, Rothlisberger and coworkers [25] investigated the LBHB model. They studied the effect of protonation pattern on the formation of ES complex in which protonation model **J** (Fig. **4**) was energetically more favourable than model **B** with a difference of 0.5 kcal mol^{-1} [25]. However, an $\Delta G^{\ddagger}$ of 36 kcal mol^{-1} was obtained for the gem-diol INTs *via* this (model **J**) protonation mechanism. This data suggests that the catalytic reaction is less favourable as this value is much higher than experimental values [125, 148]. The mechanism takes the form of Scheme **4** [152] whereby the Asp group does not share bond with the substrate (Scheme **6**). The process initiates from the introduction of substrate to the enzyme forming an ES complex going through three reaction steps to form the product complex (F′PQ).

3.2.2.1.4. Trapping Reaction Moieties Involved in HIV-1 PR—Substrate Catalysis through Crystallization: The Stepwise General Acid-Base Mechanistic Pathway

Experimental detection (with X-ray for example) of notable reaction complexes along the PES of HIV-1 PR—natural target catalysis is rarely possible due to the reactive nature of the protease [1, 129]. An exact natural substrate is cleaved by the active HIV-1 PR before the crystals grow and data collection becomes difficult [129, 170]. However, efforts were made by Hosur and coworkers [129] to present the first crystal structure of an INT structure obtained from HIV-1 PR complexion with an oligopeptide sequence consisting of eleven amino acid units (Fig. **7**). The trapped tetrahedral INT with the tethered HIV-1 PR was refined to 2.03 Å resolution. The INT was prepared through the attachment of two oxygen atoms to the scissile carbon atom of the selected undecapeptide. A non-covalent association was observed between the transient INT and the tethered HIV-1 PR [129]

These authors [129] proposed that the formation of INT and protonation of the scissile peptide nitrogen occur sequentially. A better insight on the nature and

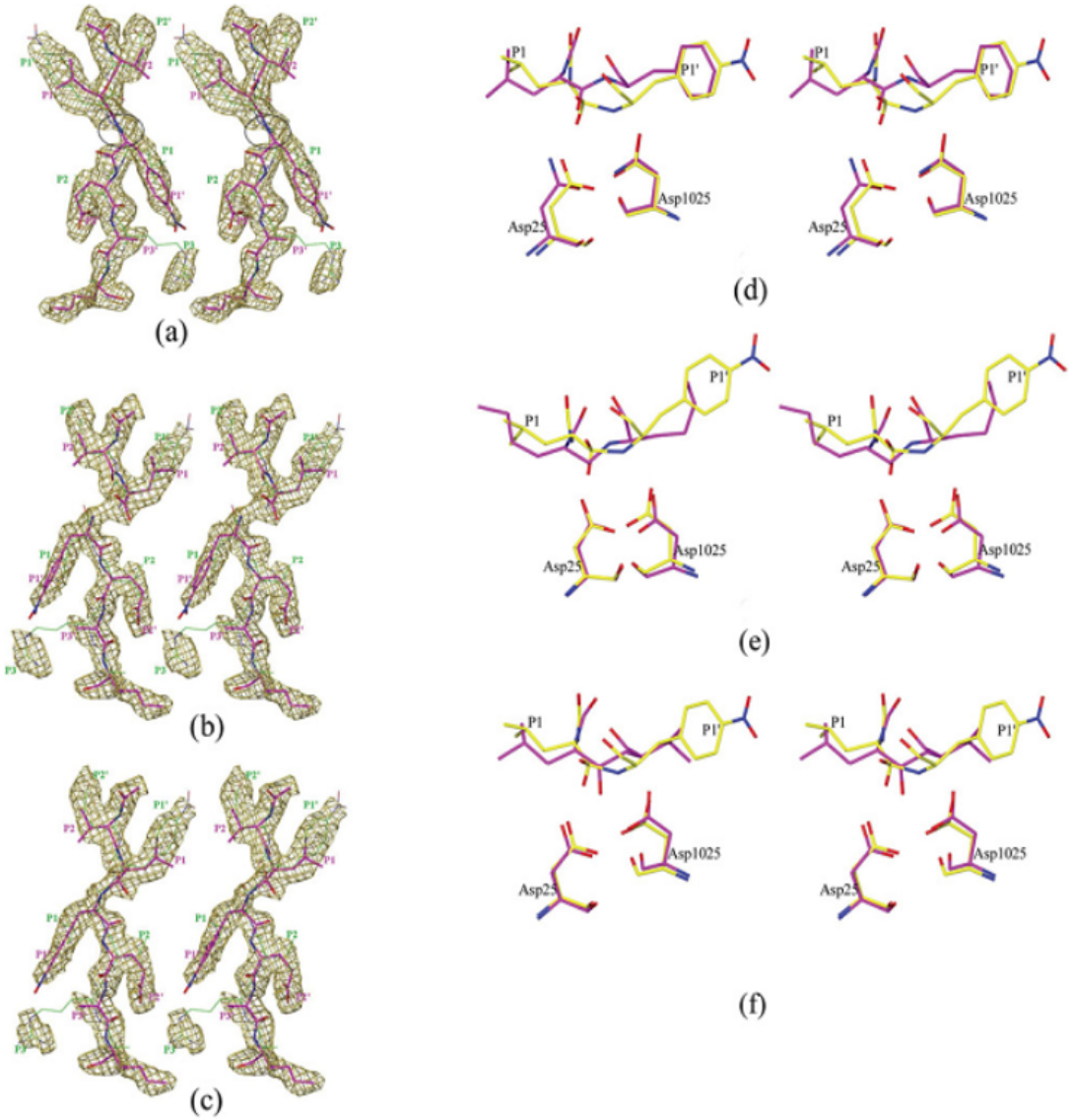


Fig. (7). The tetrahedral peptidolysis intermediate. (**a**)–(**c**) Stereo diagrams of $2F_o–F_c$ maps (contoured at 0.8σ). (**a**) The substrate is refined as a regular peptide. The region of the model between the P1 and P1′ residues, which does not fit properly in the density, is circled. (**b**) The substrate is refined as a cleaved peptide; (**c**) the substrate is refined as a reaction intermediate. Two orientations of the substrate are shown with magenta and green carbon atoms; sticks in the magenta model have been made thicker to allow easy tracing of the peptide chain. (**d**)–(**f**) Stereo diagrams of structural superposition of the reaction intermediate (yellow carbon) on to: (**d**) the regular peptide from the structure 1KJH (PDB code) (magenta carbon), (**e**) the reduced amide inhibitor MVT101 from the structure 4HVP (magenta carbon) and (**f**) the hydroxyethylene inhibitor U85548E from the structure 8HVP (magenta carbon). Only one orientation of the substrate is shown for the sake of clarity. Similar features are also present in the other orientation. Only protein Cα atoms were used for superposition [131]. Image and details are taken from literature[131] with permission "Copyright (2005) Biochemical Society Portland Press".

conformation of this INT candidate may potentially be useful in the prediction or synthesis of tighter HIV-1 PR inhibitor binders [129]. A similar observation was made from another crystallographic analysis [171]. The product segments of the HIV-1 PR—RT-RH (refined to 1.65 Å) was co-crystallized *via* standard simulated annealing (SA) procedures and the amplitude-based maximum likelihood target function [172]. The investigation also featured the observation of a LBHB between OD1 of both Asp25 and Asp25′ with an interatomic distance of 2.30 Å. Adopting Northrop [136] isomechanistic hypothesis (Scheme **6**), Das *et al.* [171] noted four hydrogen-bonding interactions between the dyad and the decapeptide substrate. From their [171] observations and discussions on the protonation pattern of the ionisable aspartates through the analysis of the

hydrogen bonds at the catalytic centre (Fig. **8**), the reaction proceeds with protonation model **J** (Fig. **4**). Within the crystal complex, the predicted separation distance between the scissile amide atoms (C----N) of the natural substrate was 2.67 Å. These two contributions [129, 171] serve as potential answer to the experimental detection of reaction complexes along HIV-1 PR—substrate PES.

Another attempt to capture reacting structures along the PES of HIV-1 PR—substrate complex was done by Bihani *et al.* [173]. The product segments of the hydrolytic action of HIV-1 PR on the matrix and capsid (MA-CA) junction (Table **1**) were elucidated with X-ray crystallography refined to 2.0 Å resolution. Using a similar approach described in literature, [129, 171] they [173] have enabled us to summarise the hydrogen bond distances within the reactive Asp dyad and the natural substrate (Table **S1**; **SI**). Two more important observations were made with respect to hydrogen bonding [173] The cleavable nitrogen atom from the heptapeptide substrate forms no hydrogen bond with OD2 of Asp25′ and the OD1 of both catalytic aspartyl group forms no LBHB as seen in previous work, [171]. These two observations motivated the hypothesis that the gem-diol INT facilitates the protonation of the scissile nitrogen of the natural target with least ascription of this phenomenal hydrogen transfer to neither catalytic Asp [173]

Successful crystallographic studies on the prediction of reacting complexes along the HIV-1 PR—substrate mechanistic pathway can be attributed to the dedicated efforts of Hosur and coworkers [129, 171, 173-176]. Using a solution of an undecapeptide substrate at pH 7, Prashar *et al.* [175] reported the prime observation of a WatC at the active site of HIV-1 PR. This was carried out *via* a crystallographic analysis, which was refined to 1.69 Å resolution. This dangling water molecule forms one H-bond with the OD2 of Asp25 and another H-bond with the OD1 of Asp25′ in the product segment. The Asp dyad was hypothesized to exhibit a protonation pattern similar to LBHB (model **J**, Fig. **4**) in which the proton is shared between the OD1 of both Asp molecules. Fig. **9** shows the proposed mechanistic route for HIV-1 PR and substrate catalysis [175] based on 3-D crystal structures from literature.

Also, the first X-ray snapshot of HIV-1 PR with a natural Gag-Pol polyprotein segment has been presented (at pH 2.5) within the reactive site of the enzyme [176]. The oligopeptide substrate corresponds to ten amino acid chain of RT-RH polypeptide subunit and trapped as a tetrahedral INT in the tethered HIV-1 PR dimer refined to 1.76 Å resolution. The investigation featured the significance of accurate protonation of the ionic Asps at the active site [176]. Detailed literature review [34, 104, 105] and analysis with respect to relevant work, [90, 177, 178] were also presented [176]. This was followed by a thorough examination of the nature of the hydrogen bond and the geometry of the coplanar Asps. In short, protonation pattern **B** (Fig. **4**) was upheld to be suitable for the tetrahedral INT description [176]. The proposed mechanism leading to product formation is summarized in Scheme **3**. One of the distinctions of this investigation [176] is the observation of a short ionic hydrogen bond (SIHB) between one of the hydroxyl oxygen of the neutral diol INT and the OD2 of Asp25′ which was a result of the asymmetric binding nature of the INT structure to the reactive Asps. Provided in Table **S1**, are preferential hydrogen bond distances within the active site of the HIV-1 PR. The SIHB presents a very sort hydrogen bond length at 2.2 Å. It was also reiterated that the separation of the INT moiety in the hydrolytic pathway

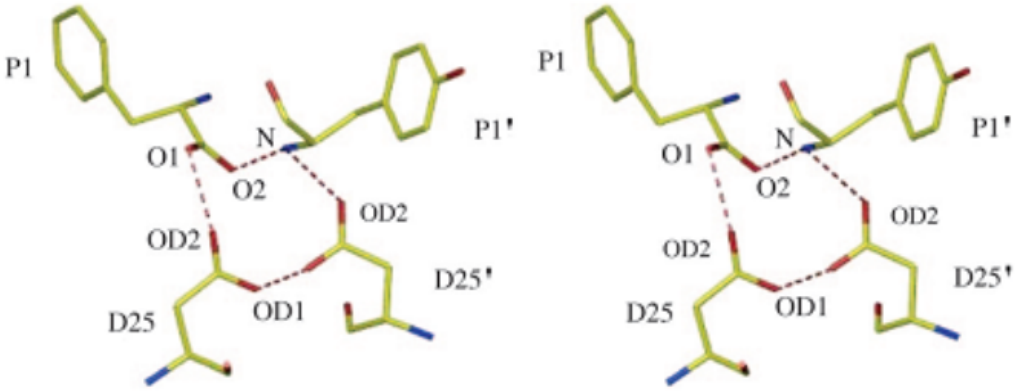


Fig. (8). Hydrogen-bonding interactions at the catalytic centre are shown by dotted lines. Picture adapted from literature[172] with permission "Copyright (2006) National Academy of Sciences, U.S.A".

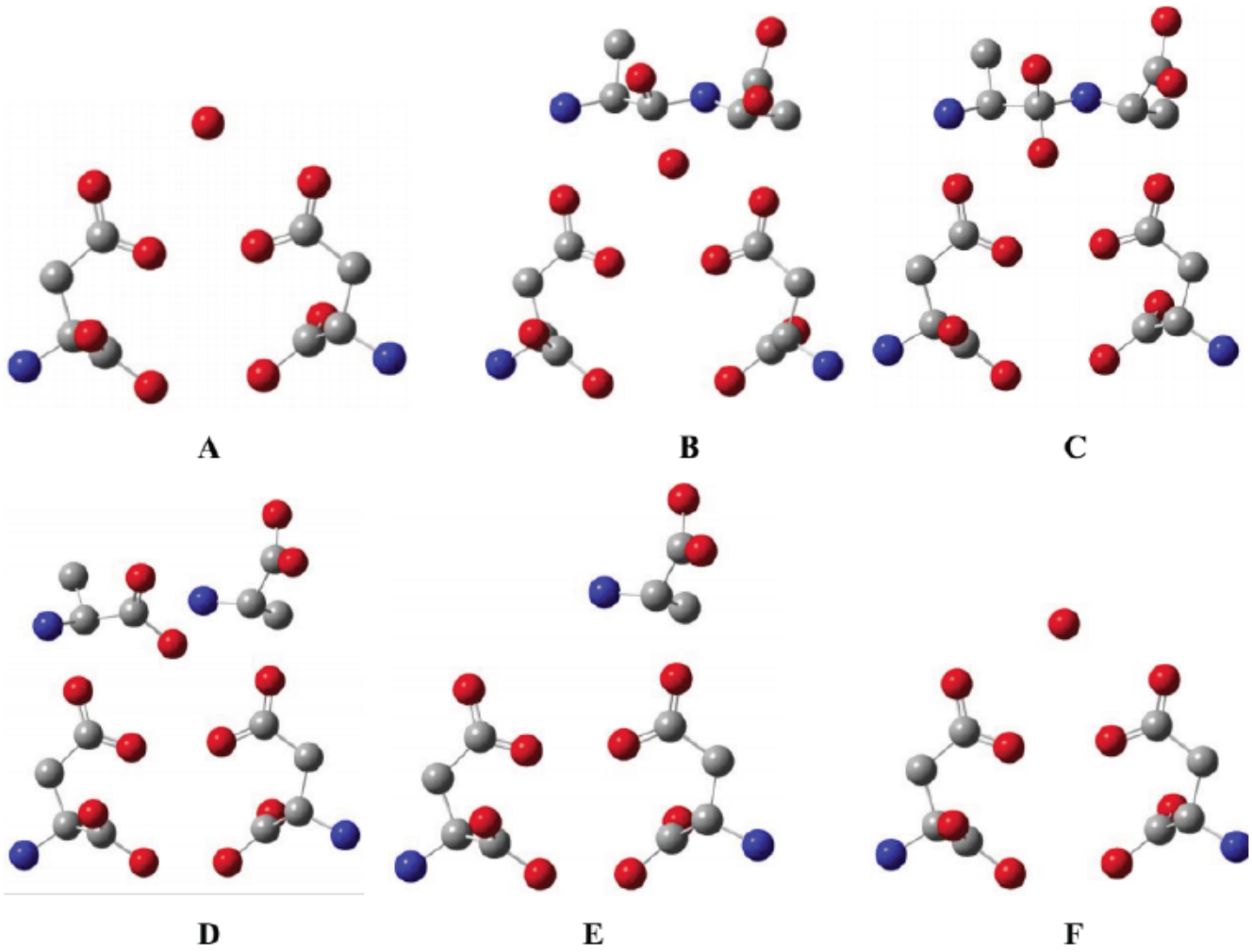


Fig. (9). Proposed sequence of steps (A–F) in the cleavage of natural substrates by HIV-1 protease redrawn from literature[176] (hydrogen atoms were omitted for clarity). Main chain atoms of substrate peptide and Asp molecules are shown in grey=carbon, blue=nitrogen and red=oxygen). Each figure is based on the structure indicated: (A) Bound WatC to ligand-free tethered HIV-1 PR Asp dyad, PDB code 1LV1.[175] (B) Modelled ES complex. (C) Formation of INT, PDB code 3MIM.[177] (D) Breaking of the scissile bond with the product peptides bound in the active site, PDB code 2NPH.[172] (E) diffusion of a unit of the separated products out of the active site and WatC into the active site, PDB code 2WHH.[176] (F) Release of the other product segment and movement of WatC into its original position, PDB code 1LV1.[175]

is the rate-determining step as observed in the literature [25, 113, 118, 176].

An autoproteolytic peptide sequence cleavage mechanism by WT, I54V, V32I and I47V HIV-1 PR mutants were investigated [152, 153]. To capture notable moieties along the reaction pathway of these drug-resistant—peptide systems, Weber and coworkers [152, 153] determined the crystal structures at high resolution (1.2–1.5 Å). In the study, the proposed mechanism of peptide bond cleavage is presented in Scheme **4** whereby the protonation is on the OD1 of Asp25′ [152]. The trapped gem-diol tetrahedral INT structures along the PRs-catalysed peptide hydrolysis reaction have no contacts with the catalytic Asps. However, the Asp group was observed to be involved in the activation of WatC and complex stability. They therefore, proposed a modified mechanism (Scheme **4**) of the substrate cleavage reaction that is devoid of the PRs forming a covalent bond as well as hydrogen transfer to the substrate. The rate-limiting step was suggested to be the decomposition of the gem-diol INT (step D; Scheme **4**) [152]

"Snapshots" of three successive steps along the reaction coordinate of V32I and I47V—nonapeptide sequence were obtained through a careful X-ray crystallographic analysis [153]. Unravelling these trapped

conformations, the metastable diol INT, the cleaving C----N product segments and the C-terminus product complex were revealed. Both WT and mutant PRs were observed to exhibit conserved interactions with the complexed peptides, [153] a related hypothesis from Schramm and coworkers [146]. A reaction mechanism was proposed with a proton placed on OD1 of Asp25′ (to initiate reactivity) in a five-step process (similar to Fig. **9**) for these enzymes [153]. The authors' hypothesis of the reaction path is enumerated: (1) More than one intermediate will be observed by simply crystallizing the protein in the presence of a peptide substrate, which implies that the energy barriers of the hydrolysis reaction of substrates by these HIV-1 PRs are similar [153]. (2) It is possible that the mechanistic route of the natural substrate hydrolysis by HIV-1 PRs does not include a single rate-determining step [153] as suggested in a number of studies [25, 113, 118, 146, 176]. (3) Majority of the interactions with bound peptides are conserved in the wild type enzyme and the mutants in agreement with the report that mutants share similar transition states to wild type PR [146]. (4) The interactions observed in these new structures [153] were mapped on the scheme (Scheme **6**) for the reaction. (5) All three intermediate stages retain the short 2.3–2.4 Å hydrogen bond, which may be an LBHB, of the hydroxyl group of the peptide intermediate with one of the catalytic aspartates, as reported in lower resolution crystal structures [129, 176]

These contributions [129, 152, 153, 171, 173, 176, 179] appear to solve a pertinent research question mentioned earlier on the determination of reaction complexes along the HIV-1 PR—substrate PES through *in vitro* methods. Capturing reaction complexes of the enzymatic process through crystallographic analyses is appreciably informative especially when different intermediate structures are discovered [179]. A clearer picture of the mechanism of enzyme catalysis is crucial to facilitate the design of new inhibitors. In addition to the crystallization methods mentioned so far, the use of theoretical/computational tools, Laue diffraction and neutron crystallography approach (to capture or see hydrogen atoms) as well as time-resolved spectroscopic analysis, could serve as complementing methods towards the provision of a comprehensive picture of HIV-1 PR reaction [179].

More recently, neutron crystallography was used to capture proton shift in HIV-1 PR—darunavir system at the reactive site of the enzyme [86]. Variations of Asp protonation states that were earlier proposed [34] were observed for the first time. The catalytic Asp dyad rarely maintains single protonation model throughout the reaction process [10, 93, 107, 112, 176, 180]. Gerlits *et al.* [86] applied neutron crystallography to detect two proton transfers at 2.0 and 2.3 Å resolutions within the acidic pH of 6.0 and 4.3, respectively. A shared proton between the OD1 of both Asps (an LBHB) in the substrate-free system moved to OD2 of Asp25 while, the proton from OD1 of Asp25′ was transferred to the OH of darunavir [86]

3.2.2.1.5. The Tetrahedral Intermediate

One of the most discussed reacting complexes along the HIV-1 PR—substrate stepwise general acid-base mechanistic pathway is the tetrahedral intermediate (INT). X-ray crystallographic analyses have buttressed the existence of the gem-diol tetrahedral INT [129, 152, 153, 176, 179] (INT1 Fig. **10**) compared to the oxyanion INT (INT2 Fig. **10**), which was proposed to be possible through a tunneling effect [136]. INT1 is a typical TS analog whose conformational mimics has served as the basis for the design of nearly a dozen US Food and Drug Administration (FDA)-approved drugs for HIV-1 PR inhibition [147, 181]. In addition to theoretical investigations [25, 119, 120, 180] attempts at exploring the nature of the INTs (both INT1 and INT2), some authors have focused their research on the tetrahedral the nature of the INTs using theoretical methods.

Carnevale *et al.* [182] carried out a comparative analytical investigation on the nature of the INTs involved in the catalysis of natural substrate segments by HIV-1 PR. They [182] remarked that the neutral gem-diol INT is highly stable compared to the ionized oxyanion INT (Fig. **10**); a similar observation noted here [25, 119, 120]. The applied model was the same as the one studied before [25] with the inclusion of the enzyme frame, counterion and explicit solvent to account for the entire system's behavior in the present model [182]. Illustrated in Fig. **10**, is the QM part in which DFT/B3LYP and MP2 were applied. The gem-hydroxyl INT was predicted to be 20–30 kcal mol^{-1} more stable than the negatively charged oxyanion INT. Even if the entropic contribution is considered, the likelihood of change in the instability of this later conformation is uncertain [182].

Garrec *et al.* [10] examined the reactivity of HIV-1 PR as a model to test the accuracy of the then latest DFT methods. The adopted protonation pattern was model **D** (Fig. **4**) while Scheme **5** was the investigated mechanistic pathway. They have suggested an explicit model that accounts for the detailed sampling of each candidate structure along the reaction coordinate [10].

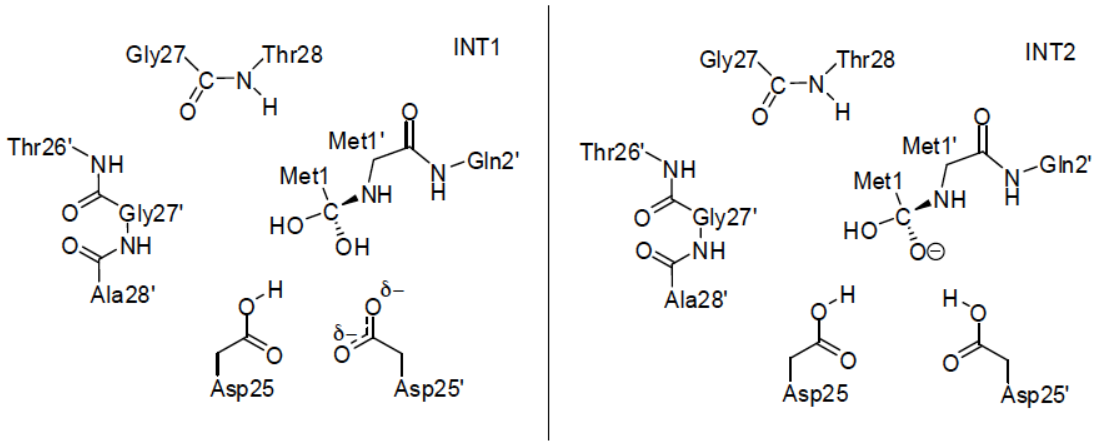


Fig. (10). Schematic representation of the INTs; INT1 is the gem-diol intermediate [25] while INT2 is the oxyanion form[112] of the substrate. Both are gas phase models in which the entire systems were partitioned into three layers. The only structural difference between the neutral and the charged reaction intermediate is the position of a proton which located on the reaction intermediate in INT1, while the both aspartic acid residues are protonated in INT2 [183].

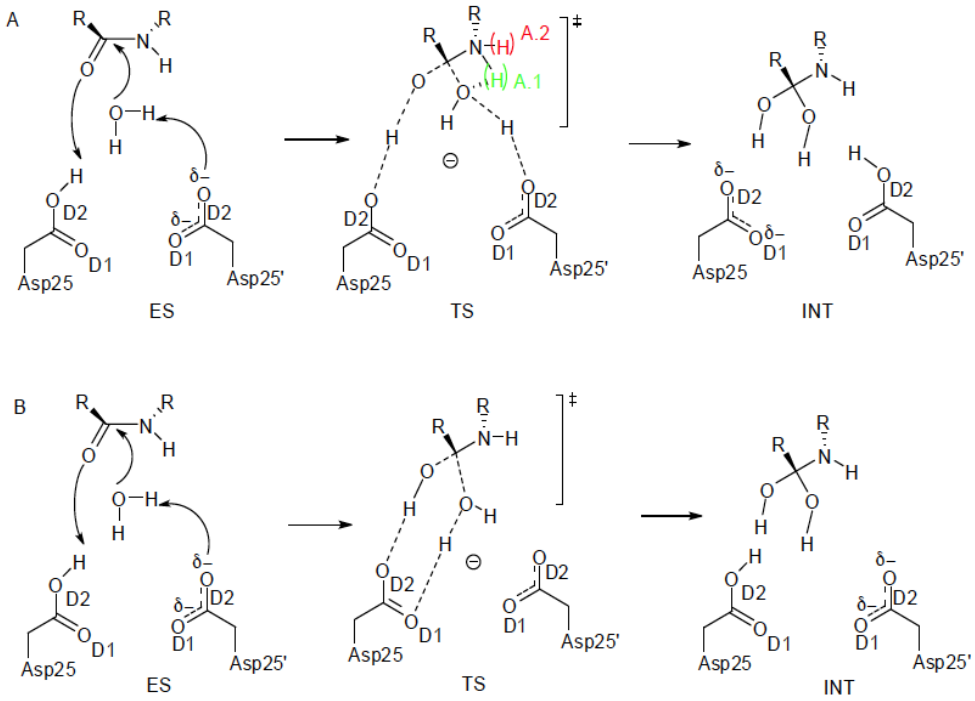
Focusing on the nature of the tetrahedral INT and the geometry of the carboxyl components of the active site Asps, these authors investigated gem-diol versus oxyanion INT and dihedral angle ^{Asp25}OD2--^{Asp25}OD1--$^{Asp25'}$OD1--$^{Asp25'}$OD2, respectively. It was noted that the oxyanion was extremely unstable while nearly all the tested DFT correlation functional yielded suitable geometry for the Asp dyad. Using a DFT/semi-empirical model through Our own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM) [183, 184] method, they [10] observed the model provides inaccurate geometry and reactivity description of the HIV-1 PR—substrate system. It appears that this inefficiency was due to the choice of an over-simplified computational model [10] involving many frozen atoms close to the active site (Fig. **S2**; **SI**).

The effect of enzymatic flexibility, individual residue contribution and conformational fluctuations leading to variations in the activation barrier energies on the PES of HIV-1 PR—p2-NC were meticulously reported by Ribeirio *et al.* [92]. The authors applied QM/MM (ONIOM) and QM/MM/MD methods to determine the nature of the barrier heights in the first step cleavage reaction of this system using protonation pattern **B** (Fig. **4**) and Scheme **7**. Their observations and conclusions are: (1) during the reaction, structural and conformational fluctuations invoked by residues' electrostatic contributions lead to the disordered energy barrier. (2) This uneven barrier distribution was denoted as "instantaneous disorder". (3) The averaged activation energy for the different conformations at 2 ns time intervals was 16.5 kcal mol^{-1} [92] and very close to the experimental value of 15.9 kcal mol^{-1} [79]. (4) Variations in activation barriers are caused by differential mechanistic routes, which the PR tends to attain. (5) One of these mechanisms involved the disappearance of Asp25 catalytic role. (6) Another catalytic route is brought by the characteristic conformational orientation of the active region [92]

The influence of frozen residues on the exploration of the PES of HIV-1 PR reaction mechanisms was recently studied [180]. The authors adopted Scheme **7** and previously reported computational protocols [92] to systematically investigate the effect of freezing different shells of residues on the activation and free energy of the first step reaction in HIV-1 PR catalysis of the substrate [180]. They observed two high $\Delta G^{\ddagger}$ in the more constrained models; 24.1 and 27.0 kcal mol^{-1} for shells of 4.00 and 5.00 Å of free residues, respectively. Freezing the atoms/residues closer to the active site often hindered the reactivity of the enzyme thereby producing higher energy barriers, [180] an observation similar to investigations from here [10, 26]. The system was observed to produce reasonable free energies when the model has a freedom of radius ≥6.00 Å away from the active site.

3.1.1.2. Concerted General Acid-Base Mechanism

The concerted general acid-base mechanism seems to be the least explored model amongst the mechanistic pathways involved in HIV-1 PR catalysis of its substrates. In 1991, Jaskólski *et al.* put forward an experimental perspective on the hydrolytic mechanism of HIV-1 PR and its substrate, [185] they have proposed a one-step catalytic model in which the nucleophilic WatC and the electrophilic aspartate (an acidic proton)



Scheme 7. Different proposed mechanism and configurations adopted by the active centre of protease redrawn from literature.[93, 181]. **A**; WatC nucleophilic attack and proton transfer to Asp25′, when it loses the planarity, the nucleophile can be more or less stabilized by the highlighted hydrogen (configurations A.1 and A.2), this is the most commonly described is mechanism in literature. **B**; WatC nucleophilic attack and proton transfer to Asp25′. Asp25 loses its catalytic role in this process and this was observed to be unfavourable.

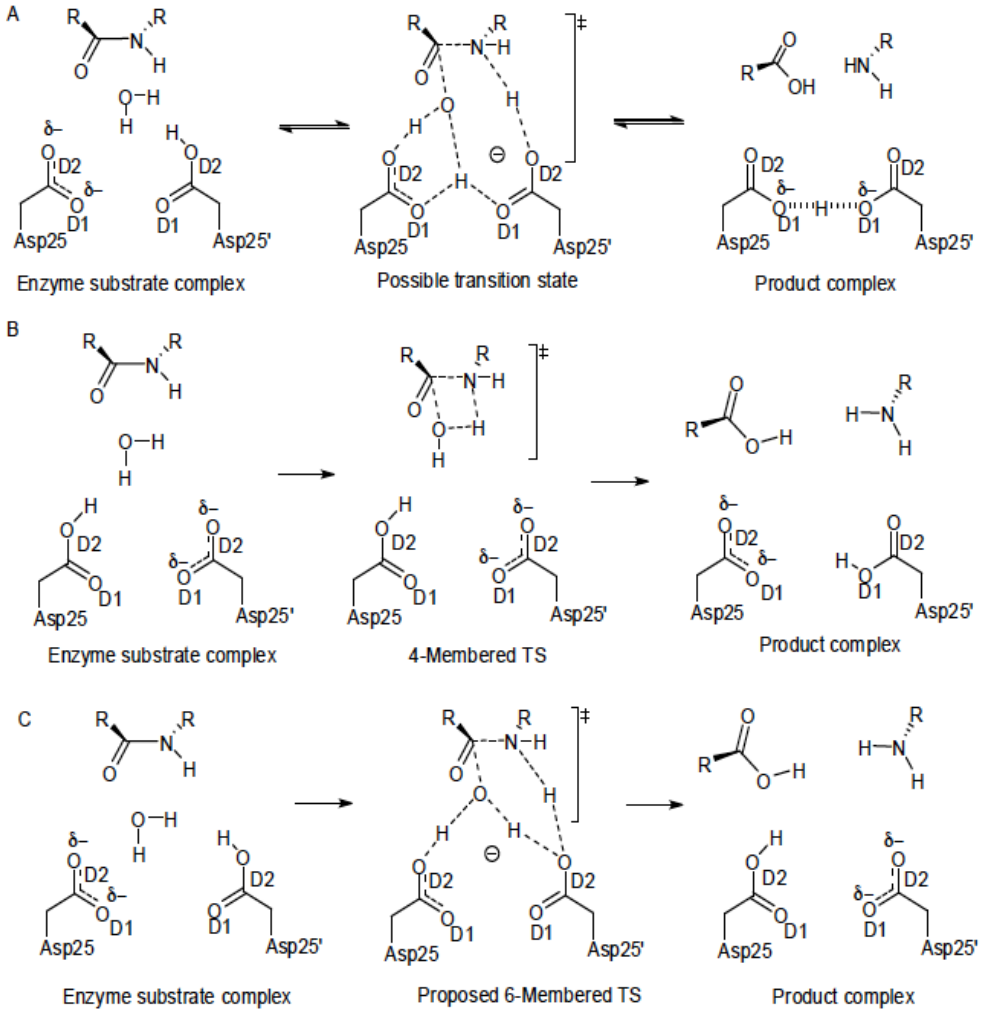
attack the scissile bond in a synchronous manner. At the beginning of the reaction, the acidic proton is located on the OD2 of the aspartate that is proximal to the N atom of the approaching amide [185]. The post-reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms [185] (model **J**; Fig. **4**) We summarize this experimental process in Scheme **8A**. It is quite puzzling to observe that only two theoretical research outputs are found with respect to this long time [185] perspective. Note that the 6-membered ring TS in Scheme **8C** is proposed by us.

In 2000, Park *et al.* [1] gave an investigation on the catalytic mechanism for Asp PRs using (MP2) [124] and Hartree-Fock (HF) [108] theories coupled with the 6-31G(2d) [109] basis set. They used a model reaction where formamide represents the substrate, while acetate and acetic acid represent the catalytic Asp residues [1]. The studied concerted general model gave an activation free energy of 30 kcal mol^{-1} and featured an 8-membered ring TS structure (Fig. S**3**; SI).

The latest attempt on the concerted general catalytic mechanism of HIV-1 with a peptide substrate was presented by Krzemińska *et al.* in 2016 [120]. The study utilizes a model in which both electrostatic and dynamic contributions were accounted for. The applied theoretical methods (umbrella sampling method; QM/MM/MD) enabled these authors [120] to explore a 4-membered ring TS model devoid of the catalytic (or general acid-base) function of the aspartic moieties (Scheme **8B**) with a characteristic high free energy barrier (43.5 kcal mol^{-1}). This theoretical observation is quite possible as reported in related chemical and enzymatic reactions [27, 163-169, 186] from our group whereby the 4-membered cyclic TS structures rarely yield better results than the 6-membered models (Scheme **8C**).

4. GENERAL OVERVIEW

In this section, some important concepts are highlighted to provide future perspectives. A careful observation of literature reveals that recognition of the substrate and the reaction mechanism of the HIV-1 PR—substrate is rather complex and the possibility of describing the process in terms of the role of water molecule(s), the Asp dyad protonation state and the rate-limiting step(s) seems uneasy. These stated conceptual parameters are crucial and interwoven when investigating the catalytic process of HIV-1 PR.



Scheme 8. The concerted general acid-base HIV-1 PR mechanism via Asp25′ OD2 protonation, (**A**) possible concerted mechanistic pathway based on Jaskólski et al. [186] hypothesis. (**B**) Concerted 4-membered ring model redrawn from literature,[122] the catalytic role of the aspartate groups seems apparently loss and (**C**) Proposed concerted 6-membered cyclic HIV-1 PR enzymatic mechanism.

Based on investigations from both theoretical and experimental, HIV-1 PR recognizes its natural substrate through; substrate modulation, substrate groove, conserved substrate shape, interdependence conformational adaptability of both enzyme and targets [17, 19, 41, 42, 72, 74, 78]. A notable contrasting opinion from computational simulations [46] is that substrate recognition of HIV-1 PR seems to be based on the conformational specificity of the protease to the Gag and Gag-Pol polyprotein pool as opposed to the popular lock-and-key or induced-fit model [46]

It is generally accepted that the enzymatic mechanism of HIV-1 PR requires WatC (Schemes **3**–**8**) therefore the nucleophilic mechanistic route (Schemes **1-2**) could be less realistic. The authenticity of the gem-diol INT (Fig. **10**, INT1) is more widely accepted than the presence of the unstable oxyanion (Fig. **10**, INT2). Apart from the hydrolytic function of WatC in the hydrolysis of the substrate peptide bond, about five additional water molecules have been highlighted to be crucial to HIV-1 PR reactivity mechanism through experimental [19, 34, 153, 187] and theoretical [93, 110, 188] techniques. The moieties along the enzyme-substrate reaction mechanism are stabilized by these five water molecules through ordered orientation of their dipoles. However, they appear not to be credited for participation in the catalytic process [93]

4.1. Stepwise Versus Concerted General Acid-Base Mechanism of HIV-1 PR

The nature of the HIV-1 PR reaction mechanism has been more often investigated as stepwise general acid-base rather than the concerted process. In the stepwise reaction pathway of HIV-1 PR—substrate showcasing

up to three TS moieties (Scheme **3**), authors have suggested that the breakdown of the gem-diol INT is the rate-determining step from both experimental [113, 146] and computational [25, 113, 118, 120, 146, 176] methods. A contrasting opinion [153] in this regard is the likelihood of the existence of more than one rate-determining step due to the possibility of more than two metastable intermediates along the reaction coordinate (Scheme **3**). Some studies [92, 180] also related the first step of the enzymatic reaction mechanism leading to INT formation as the rate-limiting step.

For the concerted general mechanism (Scheme **8**), Hyland *et al.* [125, 126] had earlier mentioned that it does not yield a discrete INT, therefore, the possibility of elucidating TS analogs necessary for inhibitor design is apparently lost. However, the concerted mechanism in which the process occurs (Scheme **8C**) *via* cyclic TS in our view should be revisited as proposed experimentally [185]. In the light of literature support for similar six-membered ring, transition structures observed theoretically, [91, 163-169] a one-step concerted chemical process appears plausible. This could potentially provide an alternative insight on the catalytic mechanism of HIV-1 PR. Such an investigation appears absent computationally and would be quite informative if the possibility of a single rate-determining step could be achieved.

FUTURE PERSPECTIVE

Studies on updates HIV-1 PR are quite diverse and require distinct update directed at some specific topics. This overview serves as an update on recent advances in understanding the recognition mode and catalytic mechanism of HIV-1 PR and its natural targets through experimental and computational techniques. Theoretical approaches play a vital role in determining the recognition pattern and enzymatic mechanism of HIV-1 PR and its natural targets, which can be expanded to other related aspartic proteases. Tremendous progress is expected in years to come through the development of advanced computer software and hardware, which would clarify the general reaction pathway for the HIV-1 PR—substrate/inhibitor complex.

Future studies on the reaction mechanism of HIV-1 PR and natural substrates should involve the application of advanced computational techniques to provide plausible answers to some unresolved perspectives. In computational investigations, a pertinent concept towards understanding the HIV-1 PR—substrate mechanism is the sufficiency of inclusive (all atoms) and excluding (just the reactive parts) models to describe the system. Based on general observations, the use of model systems for the HIV-1 PR—natural substrate mechanism [1, 10, 25, 26, 107, 118, 146] is more popular compared to an entire representation of the enzyme—substrate system [94, 110]. This is the result of researchers preferring to simplify the model initially as well as historic limitations with computer software and hardware. It therefore, seems obvious that the correlation of theoretical results with experimental data was a crucial outlook of most of these studies [1, 10, 25, 26, 107, 118, 146].

Future computational efforts should explore not only the energy values obtained but other properties such as the geometry of the system, the thermochemistry and the accuracy of the chosen theoretical tools. 3-D crystal structure of unbound and bound HIV-1 PR is largely deposited in different databases. For example, RCSB PDB [18] has almost 200 wild-type HIV-1 PR complexed with various ligands/substrate segments at a resolution as low as 1.00 Å [189]. It is informative to note that high resolution does not necessarily mean the best starting geometry for computational simulation. For instance, 3NU3.pdb [190] was refined to 1.02 Å and the active site dihedral angel (^{Asp25}OD2--^{Asp25}OD1--$^{Asp25'}$OD1--$^{Asp25'}$OD2) is 49.43º, 4HVP [137] (refined to 2.30 Å) gave 15.57º for this angle, while as large as 67º have been noted for this important dihedral [10]. Therefore, theoretical investigations on the enzymatic mechanism of HIV-1 PR—natural substrate in years to come, would likely involve the applications of sophisticated computational techniques aimed at exploring more than energetics of the system. The possibility of integrated computational algorithms which do not involve partitioning/restraining/constraining/cropped model system of the ES mechanism would likely surface in future to accurately elucidate the HIV-1 PR catalytic process on natural substrates/ligands.

The protonation pattern of the Asp dyad is still a matter of debate, a new perspective on the protonation state of the active Asps should be investigated through advanced theoretical models. Parameters such as geometric orientation, QM-based chemical properties and thermodynamics could be examined. The outcome of such studies may potentially serve as an answer to a query on adopting a unified protonation pattern for not just the natural Gag and Gag-Pol peptide segments but also, the HIV-1 PR inhibitors.

LIST OF ABBREVIATIONS

AIDS	=	Acquired Immune Deficiency Syndrome

$\Delta G^{\ddagger}$	=	Activation free energy
ANM	=	Anisotropic Network Model
AVB	=	Approximate valence bond
AMBER	=	Assisted Model Building with Energy Refinement
BSST	=	Biased sequence search threading
BACE 1	=	Beta-secretase 1
CA	=	Capsid
WatC	=	Catalytic water
Cat D	=	Cathepsin D
COBRAMM	=	Computations at Bologna Relating *Ab-initio* and Molecular Mechanics Methods
DEE	=	Dead-end elimination
DFT	=	Density functional theory
EVB	=	Empirical valence bond
ES	=	Enzyme—substrate
FlAsH	=	Fluorescein arsenical hairpin
FDA	=	Food and Drug Administration
FEP	=	Free energy perturbation
HF	=	Hartree-Fock
HIV	=	Human Immunodeficiency Virus
HTLV-1	=	Human T-cell leukaemia virus
IN	=	Integrase
INT	=	Intermediate
IRC	=	Intrinsic reaction coordinates
KIEs	=	Kinetic Isotope Effects
LBHB	=	Low barrier hydrogen bond
MA	=	Matrix
MMFF	=	Merk Molecular Force Field
MD	=	Molecular dynamics
MM	=	Molecular mechanics
MM-PBSA	=	Molecular mechanics Poisson-Boltzmann surface area
MP	=	Moller-Plesset
MDR	=	Multidrug-resistant
NMR	=	Nuclear magnetic resonance
NC	=	Nucleocapsid
ONIOM	=	Our own N-layered Integrated molecular Orbital and molecular Mechanics
PME	=	Particle Mesh Ewald
Plm II	=	Plasmepsin II
PES	=	Potential energy surface
PR	=	Protease
PDB	=	Protein Data Bank
QM	=	Quantum mechanics
REMD	=	Replica-exchange Molecular Dynamics
RT	=	Reverse transcriptase
RH	=	RNAse H
SA	=	South Africa
p1	=	Spacer peptide 1
p2	=	Spacer peptide 2
S-groove	=	Substrate-groove
TI	=	Thermodynamic integration
3-D	=	Three-dimensional
TS	=	Transition state
VDW	=	van der Waals
WT	=	Wild type

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGMENTS

We are grateful to the College of Health Sciences, UKZN, Aspen Pharmacare, Medical Research Council (MRC) and National Research Foundation (NRF), all in South Africa, for financial support. We are also grateful to the Centre for High-Performance Computing (http://www.chpc.ac.za) for computational resources.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Park, H.; Suh, J.; Lee, S. Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism. *J. Am. Chem. Soc.,* **2000**, *122*(16), 3901-3908.

[2] Canduri, F.; Teodoro, L.G.; Fadel, V.; Lorenzi, C.C.; Hial, V.; Gomes, R.A.; Neto, J.R.; de Azevedo, W.F., Jr Structure of human uropepsin at 2.45 A resolution. *Acta Crystallogr. D Biol. Crystallogr.,* **2001**, *57*(Pt 11), 1560-1570.

[3] de Azevedo, W.F., Jr; Canduri, F.; Fadel, V.; Teodoro, L.G.; Hial, V.; Gomes, R.A. Molecular model for the binary complex of uropepsin and pepstatin. *Biochem. Biophys. Res. Commun.,* **2001**, *287*(1), 277-281.

[4] Koelsch, G. BACE1 function and inhibition: implications of intervention in the amyloid pathway of Alzheimer's disease pathology. *Molecules,* **2017**, *22*(10), 1723.

[5] Aufschnaiter, A.; Kohler, V.; Büttner, S. Taking out the garbage: cathepsin D and calcineurin in neurodegeneration. *Neural Regen. Res.,* **2017**, *12*(11), 1776-1779.

[6] Aufschnaiter, A.; Habernig, L.; Kohler, V.; Diessl, J.; Carmona-Gutierrez, D.; Eisenberg, T.; Keller, W.; Büttner, S. The coordinated action of calcineurin and cathepsin D protects against α-synuclein toxicity. *Front. Mol. Neurosci.,* **2017**, *10*, 207.

[7] Canduri, F.; Ward, R.J.; de Azevedo Júnior, W.F.; Gomes, R.A.; Arni, R.K. Purification and partial characterization of cathepsin D from porcine (Sus scrofa) liver using affinity chromatography. *Biochem. Mol. Biol. Int.,* **1998**, *45*(4), 797-803.

[8] Agbowuro, A.A.; Huston, W.M.; Gamble, A.B.; Tyndall, J.D.A. Proteases and protease inhibitors in infectious diseases. *Med. Res. Rev.,* **2018**, *38*(4), 1295-1331.

[9] Silva, A.M.; Lee, A.Y.; Gulnik, S.V.; Maier, P.; Collins, J.; Bhat, T.N.; Collins, P.J.; Cachau, R.E.; Luker, K.E.; Gluzman, I.Y.; Francis, S.E.; Oksman, A.; Goldberg, D.E.; Erickson, J.W. Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium falciparum. *Proc. Natl. Acad. Sci. USA,* **1996**, *93*(19), 10034-10039.

[10] Garrec, J.; Sautet, P.; Fleurat-Lessard, P. Understanding the HIV-1 protease reactivity with DFT: what do we gain from recent functionals? *J. Phys. Chem. B,* **2011**, *115*(26), 8545-8558.

[11] Potempa, M. The triple threat of HIV-1 protease inhibitors.*The Future of HIV-1 Therapeutics*; Springer, **2015**, pp. 203-241.

[12] Lockhat, H.A.; Silva, J.R.; Alves, C.N.; Govender, T.; Lameira, J.; Maguire, G.E.; Sayed, Y.; Kruger, H.G. Binding free energy calculations of nine FDA-approved protease inhibitors against HIV-1 subtype C I36T↑T containing 100 amino acids per monomer. *Chem. Biol. Drug Des.,* **2016**, *87*(4), 487-498.

[13] Maseko, S.B.; Natarajan, S.; Sharma, V.; Bhattacharyya, N.; Govender, T.; Sayed, Y.; Maguire, G.E.; Lin, J.; Kruger, H.G. Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies. *Protein Expr. Purif.,* **2016**, *122*, 90-96.

[14] Swanstrom, R.; Wills, J. *Synthesis, assembly, and processing of viral proteins*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, **1997**.

[15] Debouck, C.; Gorniak, J.G.; Strickler, J.E.; Meek, T.D.; Metcalf, B.W.; Rosenberg, M. Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. *Proc. Natl. Acad. Sci. USA,* **1987**, *84*(24), 8903-8906.

[16] Darke, P.L.; Nutt, R.F.; Brady, S.F.; Garsky, V.M.; Ciccarone, T.M.; Leu, C.T.; Lumma, P.K.; Freidinger, R.M.; Veber, D.F.; Sigal, I.S. HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. *Biochem. Biophys. Res. Commun.,* **1988**, *156*(1), 297-303.

[17] Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C.A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.,* **2000**, *301*(5), 1207-1220.

[18] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.,* **2000**, *28*(1), 235-242.

[19] Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C.A. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure,* **2002**, *10*(3), 369-381.

[20] Lv, Z.; Chu, Y.; Wang, Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl.),* **2015**, *7*, 95-104.

[21] Dunn, B.M. Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.,* **2002**, *102*(12), 4431-4458.

[22] Mahalingam, B.; Louis, J.M.; Hung, J.; Harrison, R.W.; Weber, I.T. Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. *Proteins,* **2001**, *43*(4), 455-464.

[23] Lee, S-K.; Potempa, M.; Kolli, M.; Özen, A.; Schiffer, C.A.; Swanstrom, R. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J. Biol. Chem.,* **2012**, *287*(16), 13279-13290.

[24] Trylska, J. Computational Modelling of Protonation Equilibria and Reaction Mechanism of HIV-1 Protease., **2001**.

[25] Piana, S. Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations. *J. Phys. Chem. B,* **2004**, *108*(30), 11139-11149.

[26] Altoè, P. A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction. *Theor. Chem. Acc.,* **2007**, *118*(1), 219-240.

[27] Honarparvar, B.; Govender, T.; Maguire, G.E.; Soliman, M.E.; Kruger, H.G. Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.,* **2014**, *114*(1), 493-537.

[28] Forli, S.; Olson, A.J. Computational challenges of structure-based approaches applied to HIV. *The Future of HIV-1 Therapeutics*; Springer, **2015**, pp. 31-51.

[29] Pintro, V.O.; de Azevedo, W.F.; Filgueira, W. Optimized virtual screening workflow: Towards target-based polynomial scoring functions for HIV-1 protease. *Comb. Chem. High Throughput Screen.,* **2017**, *20*(9), 820-827.

[30] Pettersen, E.F., *et al.*, *UCSF chimera - A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.

[31] Gulnik, S.; Erickson, J.W.; Xie, D. HIV protease: enzyme function and drug resistance. *Vitam. Horm.,* **2000**, *58*, 213-256.

[32] Rodríguez-Barrios, F.; Gago, F. HIV protease inhibition: limited recent progress and advances in understanding current pitfalls. *Curr. Top. Med. Chem.,* **2004**, *4*(9), 991-1007.

[33] Nijhuis, M.; van Maarseveen, N.M.; Boucher, C.A. HIV protease resistance and viral fitness. *Curr. Opin. HIV AIDS,* **2007**, *2*(2), 108-115.

[34] Brik, A.; Wong, C-H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.,* **2003**, *1*(1), 5-14.

[35] Sussman, F.; Villaverde, M.C.; Domínguez, J.L.; Danielson, U.H. On the active site protonation state in aspartic proteases: implications for drug design. *Curr. Pharm. Des.,* **2013**, *19*(23), 4257-4275.

[36] Deo, S.K.; Lewis, J.C.; Daunert, S. Bioluminescence detection of proteolytic bond cleavage by using recombinant aequorin. *Anal. Biochem.,* **2000**, *281*(1), 87-94.

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

[38] Pawar, S.A.; Jabgunde, A.M.; Govender, P.; Maguire, G.E.; Kruger, H.G.; Parboosing, R.; Soliman, M.E.; Sayed, Y.; Dhavale, D.D.; Govender, T. Synthesis and molecular modelling studies of novel carbapeptide analogs for inhibition of HIV-1 protease. *Eur. J. Med. Chem.,* **2012**, *53*, 13-21.

[39] Makatini, M.M.; Petzold, K.; Alves, C.N.; Arvidsson, P.I.; Honarparvar, B.; Govender, P.; Govender, T.; Kruger, H.G.; Sayed, Y.; JerônimoLameira, ; Maguire, G.E.; Soliman, M.E. Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors. *J. Enzyme Inhib. Med. Chem.,* **2013**, *28*(1), 78-88.

[40] Debouck, C. The HIV-1 protease as a therapeutic target for AIDS. *AIDS Res. Hum. Retroviruses,* **1992**, *8*(2), 153-164.

[41] King, N.M.; Prabu-Jeyabalan, M.; Nalivaika, E.A.; Schiffer, C.A. Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem. Biol.,* **2004**, *11*(10), 1333-1338.

[42] Prabu-Jeyabalan, M.; Nalivaika, E.A.; Romano, K.; Schiffer, C.A. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *J. Virol.,* **2006**, *80*(7), 3607-3616.

[43] Ozer, N.; Haliloglu, T.; Schiffer, C.A. Substrate specificity in HIV-1 protease by a biased sequence search method. *Proteins,* **2006**, *64*(2), 444-456.

[44] Ozer, N.; Schiffer, C.A.; Haliloglu, T. *Predicting substrate specificity in HIV-1 protease*; Computer Aided Drug Design & Development Society in Turkey, **2006**.

[45] Ozen, A.; Haliloğlu, T.; Schiffer, C.A. Dynamics of preferential substrate recognition in HIV-1 protease: Redefining the substrate envelope. *J. Mol. Biol.,* **2011**, *410*(4), 726-744.

[46] Shen, Y.; Altman, M.D.; Ali, A.; Nalam, M.N.; Cao, H.; Rana, T.M.; Schiffer, C.A.; Tidor, B. Testing the substrate-envelope hypothesis with designed pairs of compounds. *ACS Chem. Biol.,* **2013**, *8*(11), 2433-2441.

[47] Perez, M.A.; Fernandes, P.A.; Ramos, M.J. Substrate recognition in HIV-1 protease: A computational study. *J. Phys. Chem. B,* **2010**, *114*(7), 2525-2532.

[48] Abdel-Rahman, H.M.; Al-karamany, G.S.; El-Koussi, N.A.; Youssef, A.F.; Kiso, Y. HIV protease inhibitors: peptidomimetic drugs and future perspectives. *Curr. Med. Chem.,* **2002**, *9*(21), 1905-1922.

[49] Tözsér, J.; Zahuczky, G.; Bagossi, P.; Louis, J.M.; Copeland, T.D.; Oroszlan, S.; Harrison, R.W.; Weber, I.T. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. *Eur. J. Biochem.,* **2000**, *267*(20), 6287-6295.

[50] Jones, D.T.; Tress, M.; Bryson, K.; Hadley, C. Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins,* **1999**, *37*(S3)(Suppl. 3), 104-111.

[51] Pitera, J.W.; Van Gunsteren, W.F. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simul.,* **2002**, *28*(1-2), 45-65.

[52] Shirts, M.R. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.,* **2003**, *119*(11), 5740-5761.

[53] Blondel, A. Ensemble variance in free energy calculations by thermodynamic integration: theory, optimal "Alchemical" path, and practical solutions. *J. Comput. Chem.,* **2004**, *25*(7), 985-993.

[54] Özen, A.; Haliloğlu, T.; Schiffer, C.A. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *J. Mol. Biol.,* **2011**, *410*(4), 726-744.

[55] Alder, B.J.; Wainwright, T.E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.,* **1959**, *31*(2), 459-466.

[56] Chellappan, S.; Kiran Kumar Reddy, G.S.; Ali, A.; Nalam, M.N.; Anjum, S.G.; Cao, H.; Kairys, V.; Fernandes, M.X.; Altman, M.D.; Tidor, B.; Rana, T.M.; Schiffer, C.A.; Gilson, M.K. Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis. *Chem. Biol. Drug Des.,* **2007**, *69*(5), 298-313.

[57] Altman, M.D.; Ali, A.; Reddy, G.S.; Nalam, M.N.; Anjum, S.G.; Cao, H.; Chellappan, S.; Kairys, V.; Fernandes, M.X.; Gilson, M.K.; Schiffer, C.A.; Rana, T.M.; Tidor, B. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. *J. Am. Chem. Soc.,* **2008**, *130*(19), 6099-6113.

[58] Nalam, M.N.; Schiffer, C.A. New approaches to HIV protease inhibitor drug design II: testing the substrate envelope hypothesis to avoid drug resistance and discover robust inhibitors. *Curr. Opin. HIV AIDS,* **2008**, *3*(6), 642-646.

[59] Nalam, M.N.; Ali, A.; Altman, M.D.; Reddy, G.S.; Chellappan, S.; Kairys, V.; Ozen, A.; Cao, H.; Gilson, M.K.; Tidor, B.; Rana, T.M.; Schiffer, C.A. Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.,* **2010**, *84*(10), 5368-5378.

[60] Alvizo, O.; Mittal, S.; Mayo, S.L.; Schiffer, C.A. Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease. *Protein Sci.,* **2012**, *21*(7), 1029-1041.

[61] Dunbrack, R.L., Jr; Cohen, F.E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.,* **1997**, *6*(8), 1661-1681.

[62] Pettit, S.C.; Henderson, G.J.; Schiffer, C.A.; Swanstrom, R. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J. Virol.,* **2002**, *76*(20), 10226-10233.

[63] Pettit, S.C.; Clemente, J.C.; Jeung, J.A.; Dunn, B.M.; Kaplan, A.H. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. *J. Virol.,* **2005**, *79*(16), 10601-10607.

[64] Beck, Z.Q.; Hervio, L.; Dawson, P.E.; Elder, J.H.; Madison, E.L. Identification of efficiently cleaved substrates for HIV-1 protease using a phage display library and use in inhibitor development. *Virology,* **2000**, *274*(2), 391-401.

[65] Rhee, S-Y.; Taylor, J.; Fessel, W.J.; Kaufman, D.; Towner, W.; Troia, P.; Ruane, P.; Hellinger, J.; Shirvani, V.; Zolopa, A.; Shafer, R.W. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob. Agents Chemother.,* **2010**, *54*(10), 4253-4261.

[66] Nicholls, A.; Sharp, K.A.; Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins,* **1991**, *11*(4), 281-296.

[67] Tie, Y.; Boross, P.I.; Wang, Y.F.; Gaddis, L.; Liu, F.; Chen, X.; Tozser, J.; Harrison, R.W.; Weber, I.T. Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of HIV-1 protease mutants with substrate analogs. *FEBS J.,* **2005**, *272*(20), 5265-5277.

[68] Özen, A.; Haliloğlu, T.; Schiffer, C.A. HIV-1 protease and substrate coevolution validates the substrate envelope as the substrate recognition pattern. *J. Chem. Theory Comput.,* **2012**, *8*(2)

[69] Patick, A.K.; Duran, M.; Cao, Y.; Shugarts, D.; Keller, M.R.; Mazabel, E.; Knowles, M.; Chapman, S.; Kuritzkes, D.R.; Markowitz, M. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob. Agents Chemother.,* **1998**, *42*(10), 2637-2644.

[70] Pai, V.B.; Nahata, M.C. Nelfinavir mesylate: a protease inhibitor. *Ann. Pharmacother.,* **1999**, *33*(3), 325-339.

[71] Özer, N.; Özen, A.; Schiffer, C.A.; Haliloğlu, T. Drug-resistant HIV-1 protease regains functional dynamics through cleavage site coevolution. *Evol. Appl.,* **2015**, *8*(2), 185-198.

[72] Atilgan, A.R.; Durell, S.R.; Jernigan, R.L.; Demirel, M.C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.,* **2001**, *80*(1), 505-515.

[73] Laco, G.S. HIV-1 protease substrate-groove: Role in substrate recognition and inhibitor resistance. *Biochimie,* **2015**, *118*, 90-103.

[74] Wang, Y.; Dewdney, T.G.; Liu, Z.; Reiter, S.J.; Brunzelle, J.S.; Kovari, I.A.; Kovari, L.C. Higher desolvation energy reduces molecular recognition in multi-drug resistant HIV-1 protease. *Biology (Basel),* **2012**, *1*(1), 81-93.

[75] Laco, G.S. Retroviral proteases: correlating substrate recognition with both selected and native inhibitor resistance. *J. Mol. Biochem.,* **2017**, *6*(2)

[76] Liu, Z.; Wang, Y.; Brunzelle, J.; Kovari, I.A.; Kovari, L.C. Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease. *Protein J.,* **2011**, *30*(3), 173-183.

[77] Yedidi, R.S.; Proteasa, G.; Martinez, J.L.; Vickrey, J.F.; Martin, P.D.; Wawrzak, Z.; Liu, Z.; Kovari, I.A.; Kovari, L.C. Contribution of the 80s loop of HIV-1 protease to the multidrug-resistance mechanism: crystallographic study of MDR769 HIV-1 protease variants. *Acta Crystallogr. D Biol. Crystallogr.,* **2011**, *67*(Pt 6), 524-532.

[78] Liu, Z.; Wang, Y.; Yedidi, R.S.; Dewdney, T.G.; Reiter, S.J.; Brunzelle, J.S.; Kovari, I.A.; Kovari, L.C. Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes. *Biochem. Biophys. Res. Commun.,* **2013**, *430*(3), 1022-1027.

[79] Deshmukh, L. Binding kinetics and substrate selectivity in HIV-1 protease−Gag interactions probed at atomic resolution by chemical exchange NMR. *Proceedings of the National Academy of Sciences,* **2017**.

[80] Maschera, B.; Darby, G.; Palú, G.; Wright, L.L.; Tisdale, M.; Myers, R.; Blair, E.D.; Furfine, E.S. Human immunodeficiency virus. Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. *J. Biol. Chem.,* **1996**, *271*(52), 33231-33235.

[81] Tözsér, J.; Bagossi, P.; Weber, I.T.; Copeland, T.D.; Oroszlan, S. Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases. *J. Biol. Chem.,* **1996**, *271*(12), 6781-6788.

[82] Fehér, A.; Weber, I.T.; Bagossi, P.; Boross, P.; Mahalingam, B.; Louis, J.M.; Copeland, T.D.; Torshin, I.Y.; Harrison, R.W.; Tözsér, J. Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *Eur. J. Biochem.,* **2002**, *269*(16), 4114-4120.

[83] Tözsér, J.; Gustchina, A.; Weber, I.T.; Blaha, I.; Wondrak, E.M.; Oroszlan, S. Studies on the role of the S4 substrate binding site of HIV proteinases. *FEBS Lett.,* **1991**, *279*(2), 356-360.

[84] Altman, M.D.; Nalivaika, E.A.; Prabu-Jeyabalan, M.; Schiffer, C.A.; Tidor, B. Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins,* **2008**, *70*(3), 678-694.

[85] Tripathi, A.; Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Kellogg, G.E. Complexity in modeling and understanding protonation states: computational titration of HIV-1-protease-inhibitor complexes. *Chem. Biodivers.,* **2007**, *4*(11), 2564-2577.

[86] Weber, I.T.; Waltman, M.J.; Mustyakimov, M.; Blakeley, M.P.; Keen, D.A.; Ghosh, A.K.; Langan, P.; Kovalevsky, A.Y. Joint X-ray/neutron crystallographic study of HIV-1 protease with clinical inhibitor amprenavir: insights for drug design. *J. Med. Chem.,* **2013**, *56*(13), 5631-5635.

[87] Gerlits, O.; Wymore, T.; Das, A.; Shen, C.H.; Parks, J.M.; Smith, J.C.; Weiss, K.L.; Keen, D.A.; Blakeley, M.P.; Louis, J.M.; Langan, P.; Weber, I.T.; Kovalevsky, A. Long-range electrostatics-induced two-proton transfer captured by neutron crystallography in an enzyme catalytic site. *Angew. Chem. Int. Ed. Engl.,* **2016**, *55*(16), 4924-4927.

[88] Yamazaki, T. NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.,* **1994**, *116*(23), 10791-10792.

[89] Smith, R.; Brereton, I.M.; Chai, R.Y.; Kent, S.B. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.,* **1996**, *3*(11), 946-950.

[90] Wang, Y.X.; Freedberg, D.I.; Yamazaki, T.; Wingfield, P.T.; Stahl, S.J.; Kaufman, J.D.; Kiso, Y.; Torchia, D.A. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry,* **1996**, *35*(31), 9945-9950.

[91] Yu, N.; Hayik, S.A.; Wang, B.; Liao, N.; Reynolds, C.H.; Merz, K.M., Jr. Assigning the protonation states of the key aspartates in β-Secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.,* **2006**, *2*(4), 1057-1069.

[92] Makatini, M.M.; Petzold, K.; Sriharsha, S.N.; Ndlovu, N.; Soliman, M.E.; Honarparvar, B.; Parboosing, R.; Naidoo, A.; Arvidsson, P.I.; Sayed, Y.; Govender, P.; Maguire, G.E.; Kruger, H.G.; Govender, T. Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. *Eur. J. Med. Chem.,* **2011**, *46*(9), 3976-3985.

[93] Ribeiro, A.J.M. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catal.,* **2015**, *5*(9), 5617-5626.

[94] Okimoto, N. Molecular Dynamics Study of HIV-1 Protease− Substrate Complex: Roles of the Water Molecules at the Loop Structures of the Active Site. *J. Am. Chem. Soc.,* **2000**, *122*(23), 5613-5622.

[95] Okimoto, N. Molecular dynamics simulations of a complex of HIV-1 protease and substrate: substrate-dependent efficiency of catalytic activity. *J. Mol. Struct. THEOCHEM,* **2001**, *543*(1), 53-63.

[96] Itoh, S.G.; Damjanovic, A.; Brooks, B.R. pH Replica-Exchange Method Based on Discrete Protonation States. *Biophys. J.,* **2012**, *102*(3), 40a-40a.

[97] Li, H.; Robertson, A.D.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins,* **2005**, *61*(4), 704-721.

[98] Dolinsky, T.J. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations., Nucleic Acids Research, 2007. 35(suppl_2): p. W522-W525

[99] Gordon, J.C. Nucleic Acids Research, 2005. 33(suppl_2): p. W368-W371

[100] McGee, T.D., Jr; Edwards, J.; Roitberg, A.E. pH-REMD simulations indicate that the catalytic aspartates of HIV-1 protease exist primarily in a monoprotonated state. *J. Phys. Chem. B,* **2014**, *118*(44), 12577-12585.

[101] Piana, S.; Carloni, P. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins,* **2000**, *39*(1), 26-36.

[102] Dodson, G.; Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.,* **1998**, *23*(9), 347-352.

[103] Makatini, M.M.; Petzold, K.; Sriharsha, S.N.; Soliman, M.E.; Honarparvar, B.; Arvidsson, P.I.; Sayed, Y.; Govender, P.; Maguire, G.E.; Kruger, H.G.; Govender, T. Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease. *Bioorg. Med. Chem. Lett.,* **2011**, *21*(8), 2274-2277.

[104] Meek, T.D.; Dayton, B.D.; Metcalf, B.W.; Dreyer, G.B.; Strickler, J.E.; Gorniak, J.G.; Rosenberg, M.; Moore, M.L.; Magaard, V.W.; Debouck, C. Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. *Proc. Natl. Acad. Sci. USA,* **1989**, *86*(6), 1841-1845.

[105] Smith, R.; Brereton, I.M.; Chai, R.Y.; Kent, S.B. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.,* **1996**, *3*(11), 946-950.

[106] Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.,* **2001**, *123*(36), 8730-8737.

[107] Piana, S.; Carloni, P.; Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.,* **2002**, *11*(10), 2393-2402.

[108] Piana, S.; Carloni, P.; Parrinello, M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.,* **2002**, *319*(2), 567-583.

[109] Becke, A.D. A new mixing of Hartree–Fock and local density‐functional theories. *J. Chem. Phys.,* **1993**, *98*(2), 1372-1377.

[110] Rassolov, V.A. 6‐31G* basis set for third‐row atoms. *J. Comput. Chem.,* **2001**, *22*(9), 976-984.

[111] Bjelic, S.; Aqvist, J. Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry,* **2006**, *45*(25), 7709-7723.

[112] Chatfield, D.C.; Brooks, B.R. HIV-1 protease cleavage mechanism elucidated with molecular-dynamics simulation. *J. Am. Chem. Soc.,* **1995**, *117*(20), 5561-5572.

[113] Trylska, J.; Grochowski, P.; McCammon, J.A. The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein Sci.,* **2004**, *13*(2), 513-528.

[114] Rodriguez, E.J.; Angeles, T.S.; Meek, T.D. Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus 1 protease. *Biochemistry,* **1993**, *32*(46), 12380-12385.

[115] Lee, H.; Darden, T.A.; Pedersen, L.G. An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease. *J. Am. Chem. Soc.,* **1996**, *118*(16), 3946-3950.

[116] Liu, H.; Müller-Plathe, F.; van Gunsteren, W.F. A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *J. Mol. Biol.,* **1996**, *261*(3), 454-469.

[117] Silva, A.M.; Cachau, R.E.; Sham, H.L.; Erickson, J.W. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *J. Mol. Biol.,* **1996**, *255*(2), 321-346.

[118] Okimoto, N. Gag protein hydrolysis mechanism by HIV-1 protease - Investigation by semiempirical molecular orbital method. *Nippon Kagaku Kaishi,* **1997**, (4), 260-266.

[119] Okimoto, N. Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method. *J. Am. Chem. Soc.,* **1999**, *121*(32), 7349-7354.

[120] Trylska, J.; Bała, P.; Geller, M.; Grochowski, P. Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. *Biophys. J.,* **2002**, *83*(2), 794-807.

[121] Krzemińska, A.; Moliner, V.; Świderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J. Am. Chem. Soc.,* **2016**, *138*(50), 16283-16298.

[122] Grochowski, P. Density functional based parametrization of a valence bond method and its applications in quantum‐classical molecular dynamics simulations of enzymatic reactions. *Int. J. Quantum Chem.,* **1996**, *60*(6), 1143-1164.

[123] Trylska, J.; Grochowski, P.; Geller, M. Parameterization of the approximate valence bond (AVB) method to describe potential energy surface in the reaction catalyzed by HIV-1 protease. *Int. J. Quantum Chem.,* **2001**, *82*(2), 86-103.

[124] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.,* **1964**, *136*(3B), B864.

[125] Møller, C.; Plesset, M.S. Note on an approximation treatment for many-electron systems. *Phys. Rev.,* **1934**, *46*(7), 618.

[126] Hyland, L.J.; Tomaszek, T.A., Jr; Roberts, G.D.; Carr, S.A.; Magaard, V.W.; Bryan, H.L.; Fakhoury, S.A.; Moore, M.L.; Minnich, M.D.; Culp, J.S. Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange. *Biochemistry,* **1991**, *30*(34), 8441-8453.

[127] Hyland, L.J.; Tomaszek, T.A., Jr; Meek, T.D. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry,* **1991**, *30*(34), 8454-8463.

[128] Antonov, V.K.; Ginodman, L.M.; Kapitannikov, Y.V.; Barshevskaya, T.N.; Gurova, A.G.; Rumsh, L.D. Mechanism of pepsin catalysis: general base catalysis by the active-site carboxylate ion. *FEBS Lett.,* **1978**, *88*(1), 87-90.

[129] Antonov, V.K.; Ginodman, L.M.; Rumsh, L.D.; Kapitannikov, Y.V.; Barshevskaya, T.N.; Yavashev, L.P.; Gurova, A.G.; Volkova, L.I. Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange. *Eur. J. Biochem.,* **1981**, *117*(1), 195-200.

[130] Kumar, M.; Prashar, V.; Mahale, S.; Hosur, M.V. Observation of a tetrahedral reaction intermediate in the HIV-1 protease-substrate complex. *Biochem. J.,* **2005**, *389*(Pt 2), 365-371.

[131] Torrie, G.M.; Valleau, J.P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.,* **1974**, *28*(4), 578-581.

[132] Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* **2011**, *1*(6), 932-942.

[133] Laio, A.; VandeVondele, J.; Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.,* **2002**, *116*(16), 6941-6947.

[134] Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.,* **1988**, *38*(6), 3098-3100.

[135] Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter,* **1988**, *37*(2), 785-789.

[136] Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.,* **1996**, *6*(1), 15-50.

[137] Northrop, D.B. Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.,* **2001**, *34*(10), 790-797.

[138] Miller, M.; Schneider, J.; Sathyanarayana, B.K.; Toth, M.V.; Marshall, G.R.; Clawson, L.; Selk, L.; Kent, S.B.; Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science,* **1989**, *246*(4934), 1149-1152.

[139] Schock, H.B.; Garsky, V.M.; Kuo, L.C. Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. *J. Biol. Chem.,* **1996**, *271*(50), 31957-31963.

[140] Aqvist, J.; Warshel, A. Simulation of Enzyme-Reactions using Valence-Bond Force-Fields and other Hybrid Quantum-Classical Approaches. *Chem. Rev.,* **1993**, *93*(7), 2523-2544.

[141] Kamerlin, S.C.L.; Warshel, A. The empirical valence bond model: theory and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.,* **2011**, *1*(1), 30-45.

[142] Warshel, A. Computer modeling of chemical reactions in enzymes and solutions; Wiley New York, **1991**.

[143] Bitencourt-Ferreira, G.; de Azevedo, W.F. Development of a machine-learning model to predict Gibbs free energy of binding for protein-ligand complexes. *Biophys. Chem.,* **2018**, *240*, 63-69.

[144] Lima, A.N.; Philot, E.A.; Trossini, G.H.; Scott, L.P.; Maltarollo, V.G.; Honorio, K.M. Use of machine learning approaches for novel drug discovery. *Expert Opin. Drug Discov.,* **2016**, *11*(3), 225-239.

[145] Xavier, M.M.; Heck, G.S.; Avila, M.B.; Levin, N.M.B.; Pintro, V.O.; Carvalho, N.L.; Azevedo, W.F. SAnDReS a computational tool for statistical analysis of docking results and development of scoring functions. *Comb. Chem. High Throughput Screen.,* **2016**, *19*(10), 801-812.

[146] Cascella, M.; Micheletti, C.; Rothlisberger, U.; Carloni, P. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *J. Am. Chem. Soc.,* **2005**, *127*(11), 3734-3742.

[147] Kipp, D.R.; Hirschi, J.S.; Wakata, A.; Goldstein, H.; Schramm, V.L. Transition states of native and drug-resistant HIV-1 protease are the same. *Proc. Natl. Acad. Sci. USA,* **2012**, *109*(17), 6543-6548.

[148] Schramm, V.L. Transition States, analogues, and drug development. *ACS Chem. Biol.,* **2013**, *8*(1), 71-81.

[149] Kipp, D.R.; Silva, R.G.; Schramm, V.L. Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease. *J. Am. Chem. Soc.,* **2011**, *133*(48), 19358-19361.

[150] Torbeev, V.Y.; Raghuraman, H.; Hamelberg, D.; Tonelli, M.; Westler, W.M.; Perozo, E.; Kent, S.B. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. *Proc. Natl. Acad. Sci. USA,* **2011**, *108*(52), 20982-20987.

[151] Polgár, L.; Szeltner, Z.; Boros, I. Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease. *Biochemistry,* **1994**, *33*(31), 9351-9357.

[152] Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.,* **1988**, *38*(6), 3098-3100.

[153] Kovalevsky, A.Y.; Chumanevich, A.A.; Liu, F.; Louis, J.M.; Weber, I.T. Caught in the Act: the 1.5 A resolution crystal structures of the HIV-1 protease and the I54V mutant reveal a tetrahedral reaction intermediate. *Biochemistry,* **2007**, *46*(51), 14854-14864.

[154] Shen, C.H.; Tie, Y.; Yu, X.; Wang, Y.F.; Kovalevsky, A.Y.; Harrison, R.W.; Weber, I.T. Capturing the reaction pathway in near-atomic-resolution crystal structures of HIV-1 protease. *Biochemistry,* **2012**, *51*(39), 7726-7732.

[155] Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure,* **1999**, *7*(3), R55-R60.

[156] Godbout, N. Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chem.,* **1992**, *70*(2), 560-571.

[157] Binkley, J.S.; Pople, J.A.; Hehre, W.J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.,* **1980**, *102*(3), 939-947.

[158] Gordon, M.S. Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J. Am. Chem. Soc.,* **1982**, *104*(10), 2797-2803.

[159] Hehre, W.J.; Stewart, R.F.; Pople, J.A. self-consistent molecular-orbital methods. i. use of gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.,* **1969**, *51*(6), 2657-2664.

[160] Collins, J.B. Self-consistent molecular orbital methods. XVII. Geometries and binding energies of second-row molecules. A comparison of three basis sets. *J. Chem. Phys.,* **1976**, *64*(12), 5142-5151.

[161] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.,* **2003**, *24*(16), 1999-2012.

[162] Gerritzen, D. Limba, and H.-H., NMR Spectroscopy. *J. Am. Chem. Soc.,* **1984**, *106*(4), 869-879.

[163] Cleland, W.W. Low-barrier hydrogen bonds and enzymatic catalysis. *Arch. Biochem. Biophys.,* **2000**, *382*(1), 1-5.

[164] Kruger, H.G. Ab initio mechanistic study of the protection of alcohols and amines with anhydrides. *J. Mol. Struct. THEOCHEM,* **2002**, *577*(2), 281-285.

[165] Gokul, V. An ab initio mechanistic understanding of the regioselective acetylation of 8,11-dihydroxy-pentacyclo[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam. *J. Mol. Struct. THEOCHEM,* **2004**, *672*(1-3), 119-125.

[166] Kruger, H.G. Experimental and computational studies of the regioselective protection of hydantoins using anhydride. *J. Mol. Struct. THEOCHEM,* **2006**, *771*(1-3), 165-170.

[167] Singh, T. Theoretical study on the formation of a pentacyclo-undecane cage lactam. *Comput. Theor. Chem.,* **2012**, *986*, 63-70.

[168] Md Abdur Rauf, S.; Arvidsson, P.I.; Albericio, F.; Govender, T.; Maguire, G.E.; Kruger, H.G.; Honarparvar, B. The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study. *Org. Biomol. Chem.,* **2015**, *13*(39), 9993-10006.

[169] Lawal, M.M.; Govender, T.; Maguire, G.E.; Honarparvar, B.; Kruger, H.G. Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study. *J. Mol. Model.,* **2016**, *22*(10), 235.

[170] Fakhar, Z. Computational model for the acylation step of the β-lactam ring: Potential application for l, d-transpeptidase 2 in mycobacterium tuberculosis. *J. Mol. Struct.,* **2017**, *1128*, 94-102.

[171] Rose, R.B.; Craik, C.S.; Douglas, N.L.; Stroud, R.M. Three-dimensional structures of HIV-1 and SIV protease

product complexes. *Biochemistry,* **1996**, *35*(39), 12933-12944.

[172] Das, A.; Prashar, V.; Mahale, S.; Serre, L.; Ferrer, J.L.; Hosur, M.V. Crystal structure of HIV-1 protease in situ product complex and observation of a low-barrier hydrogen bond between catalytic aspartates. *Proc. Natl. Acad. Sci. USA,* **2006**, *103*(49), 18464-18469.

[173] Brünger, A.T.; Adams, P.D.; Clore, G.M.; DeLano, W.L.; Gros, P.; Grosse-Kunstleve, R.W.; Jiang, J.S.; Kuszewski, J.; Nilges, M.; Pannu, N.S.; Read, R.J.; Rice, L.M.; Simonson, T.; Warren, G.L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.,* **1998**, *54*(Pt 5), 905-921.

[174] Bihani, S.; Das, A.; Prashar, V.; Ferrer, J.L.; Hosur, M.V. X-ray structure of HIV-1 protease in situ product complex. *Proteins,* **2009**, *74*(3), 594-602.

[175] Kumar, M.; Kannan, K.K.; Hosur, M.V.; Bhavesh, N.S.; Chatterjee, A.; Mittal, R.; Hosur, R.V. Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. *Biochem. Biophys. Res. Commun.,* **2002**, *294*(2), 395-401.

[176] Prashar, V.; Bihani, S.; Das, A.; Ferrer, J.L.; Hosur, M. Catalytic water co-existing with a product peptide in the active site of HIV-1 protease revealed by X-ray structure analysis. *PLoS One,* **2009**, *4*(11), e7860.

[177] Das, A.; Mahale, S.; Prashar, V.; Bihani, S.; Ferrer, J.L.; Hosur, M.V. X-ray snapshot of HIV-1 protease in action: observation of tetrahedral intermediate and short ionic hydrogen bond SIHB with catalytic aspartate. *J. Am. Chem. Soc.,* **2010**, *132*(18), 6366-6373.

[178] Ichikawa, M. The C–O vs O–H length correlation in hydrogen-bonded carboxyl groups. *Acta Crystallogr. B,* **1979**, *35*(5), 1300-1301.

[179] Yu, N.; Hayik, S.A.; Wang, B.; Liao, N.; Reynolds, C.H.; Merz, K.M., Jr Assigning the protonation states of the key aspartates in β-Secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.,* **2006**, *2*(4), 1057-1069.

[180] Weber, I.T. Reaction Intermediates Discovered in Crystal Structures of Enzymes, in Structural and Mechanistic Enzymology: Bringing Together Experiments and Computing; Christov, C.; KarabenchevaChristova, T., Eds.; **2012**, pp. 57-86.

[181] Calixto, A.R.; Ramos, M.J.; Fernandes, P.A. Influence of frozen residues on the exploration of the PES of enzyme reaction mechanisms. *J. Chem. Theory Comput.,* **2017**, *13*(11), 5486-5495.

[182] Mahalingam, A.K.; Axelsson, L.; Ekegren, J.K.; Wannberg, J.; Kihlström, J.; Unge, T.; Wallberg, H.; Samuelsson, B.; Larhed, M.; Hallberg, A. HIV-1 protease inhibitors with a transition-state mimic comprising a tertiary alcohol: Improved antiviral activity in cells. *J. Med. Chem.,* **2010**, *53*(2), 607-615.

[183] Carnevale, V. On the nature of the reaction intermediate in the HIV-1 protease: a quantum chemical study. *Comput. Phys. Commun.,* **2008**, *179*(1-3), 120-123.

[184] Svensson, M. ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels− Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition. *J. Phys. Chem.,* **1996**, *100*(50), 19357-19363.

[185] Chung, L.W.; Sameera, W.M.; Ramozzi, R.; Page, A.J.; Hatanaka, M.; Petrova, G.P.; Harris, T.V.; Li, X.; Ke, Z.; Liu, F.; Li, H.B.; Ding, L.; Morokuma, K. The ONIOM method and its applications. *Chem. Rev.,* **2015**, *115*(12), 5678-5796.

[186] Jaskólski, M.; Tomasselli, A.G.; Sawyer, T.K.; Staples, D.G.; Heinrikson, R.L.; Schneider, J.; Kent, S.B.; Wlodawer, A. Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry,* **1991**, *30*(6), 1600-1609.

[187] Lawal, M.M. DFT study of the acid‐catalyzed esterification reaction mechanism of methanol with carboxylic acid and its halide derivatives. *Int. J. Quantum Chem.,* **2017**.

[188] Wang, Y-X. Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. *J. Am. Chem. Soc.,* **1996**, *118*(49), 12287-12290.

[189] Li, Z.; Lazaridis, T. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.,* **2003**, *125*(22), 6636-6637.

[190] Ghosh, A.K.; Gemma, S.; Baldridge, A.; Wang, Y.F.; Kovalevsky, A.Y.; Koh, Y.; Weber, I.T.; Mitsuya, H. Flexible cyclic ethers/polyethers as novel P2-ligands for HIV-1 protease inhibitors: Design, synthesis, biological evaluation, and protein-ligand X-ray studies. *J. Med. Chem.,* **2008**, *51*(19), 6021-6033.

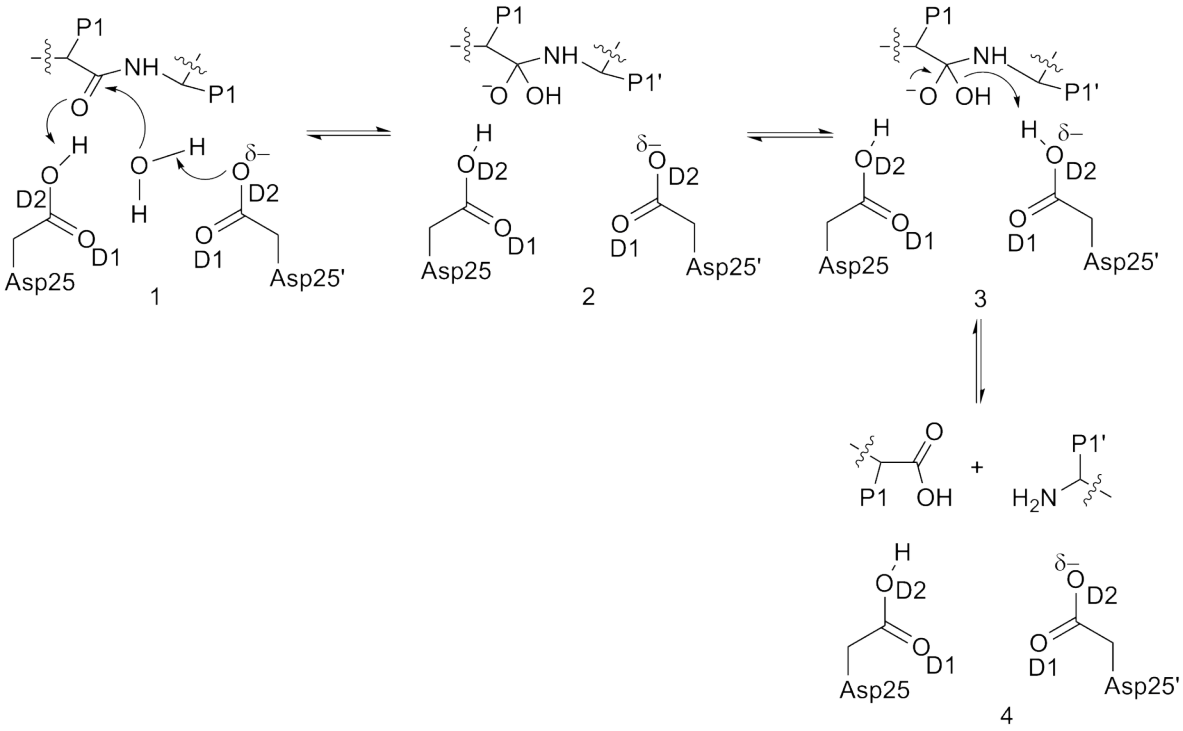
[191] Shen, C.H.; Wang, Y.F.; Kovalevsky, A.Y.; Harrison, R.W.; Weber, I.T. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS J.,* **2010**, *277*(18), 3699-3714.

Supplementary Materials

From Recognition to Reaction Mechanism: An Overview on the Interactions between HIV-1 Protease and its Natural Targets

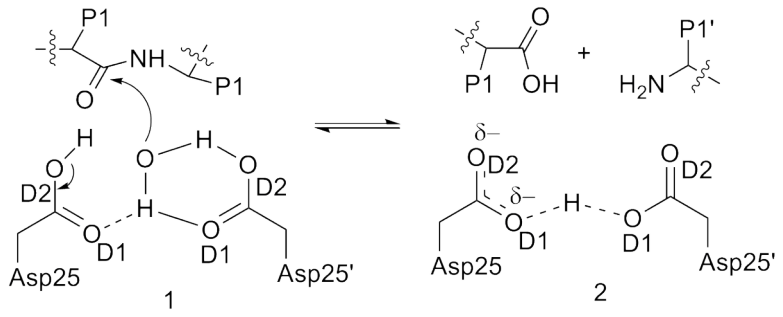
Monsurat M. Lawal[a], Zainab K. Sanusi[a], Thavendran Govender[a], Glenn E.M. Maguire[a,b], Bahareh Honarparvar[a], and Hendrik G. Kruger[a,*]

[a]*Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa;* [b]*School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa*

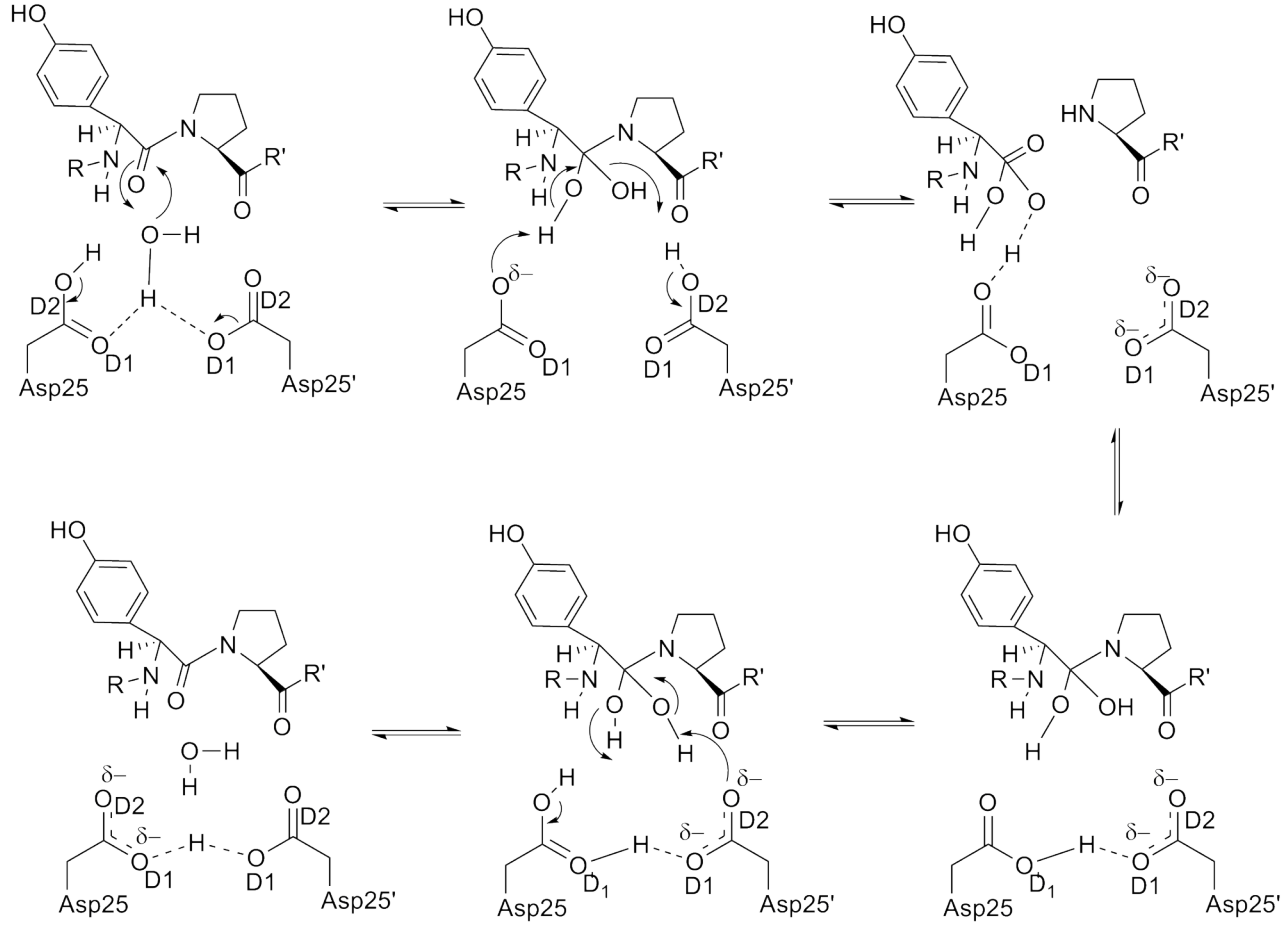


Scheme S1: Proposed stepwise catalytic mechanism for HIV-1 PR through protonation of OD2 in Asp. Complex 1 involves the transfer of electronic charges from the deionized Asp25 to the water molecule and subsequent attack of this hydrolysis water unit on the scissile carbonyl of the substrate. Compound 2 is a typical oxyanion intermediate while 3 involve the breakdown of this tetrahedral intermediate. Products are completely separated in 4 and thus released into the HIV life cycle.

0929-8673/19 $58.00+.00



Scheme S2: Proposed concerted HIV-1 PR enzymatic mechanism. Bond breaking and forming process may occur concurrently (1) and products (2) are formed.



Scheme S3: Proposed mechanism for HIV PR catalyzed incorporation of $_{18}O$ and H_2O^{18} into the peptide substrate.

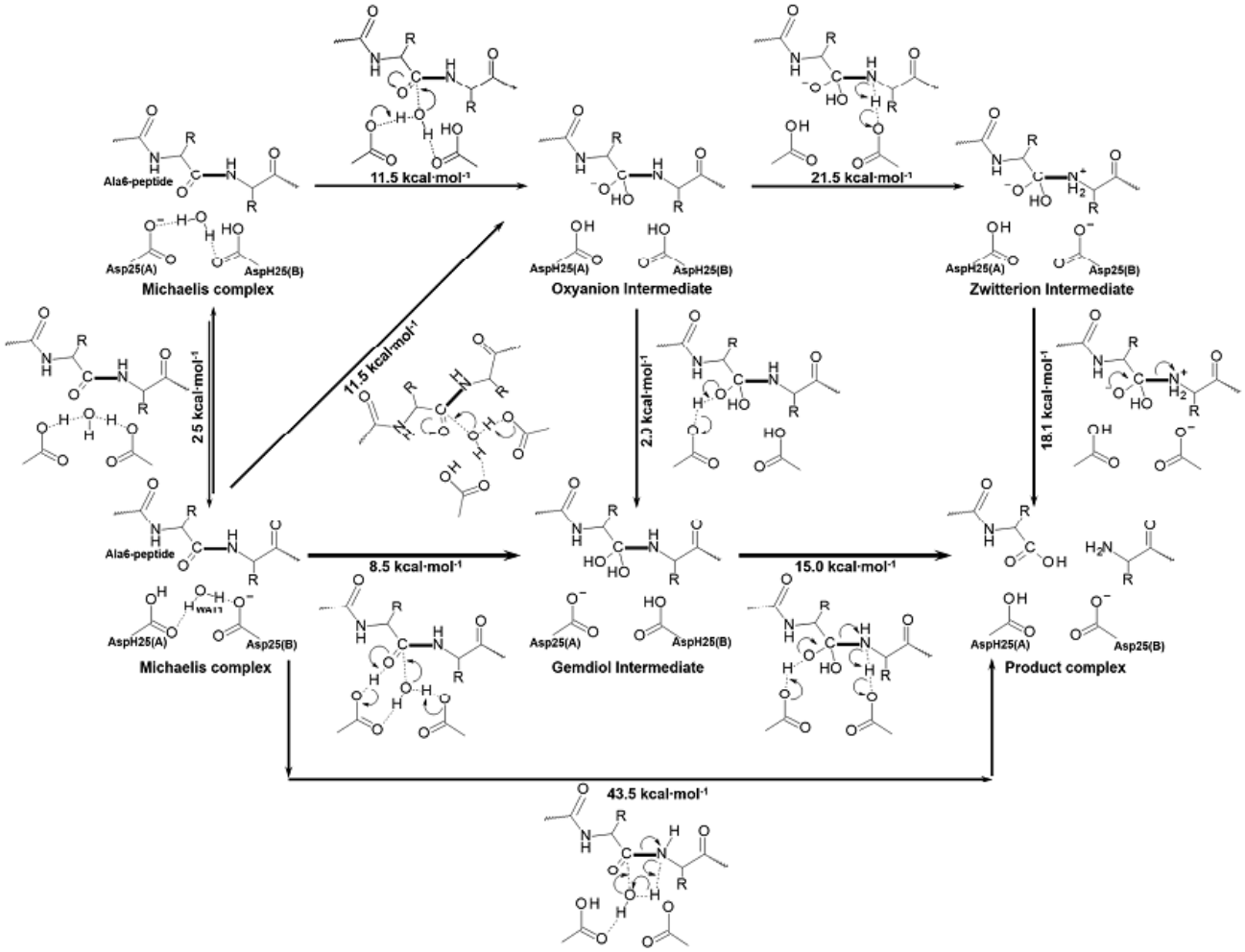


Figure S1: Schematic representation of the overall studied mechanisms for HIV-1 PR catalysed reaction of hexapeptide bond cleavage. Activation free energies, in relation to the initial reactant state, were derived from the AM1/MM PMFs corrected at M06-2X:AM1/MM level. All values are reported in kcal·mol-1. "Adapted with permission from article.[1] Copyright (2016) American Chemical Society."

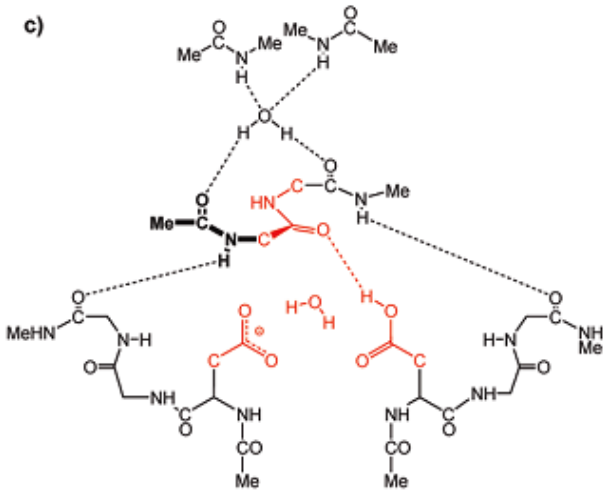


Figure S2: In ONIOM calculations, atoms in red are described at the high level.[2]

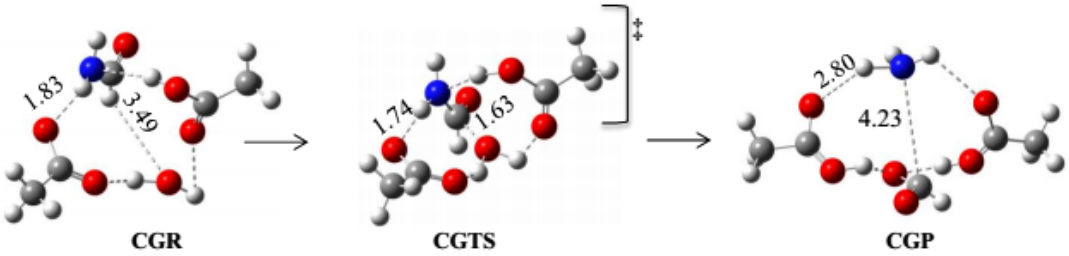


Figure S3: The concerted general acid-base mechanism involving 8-membered ring TS.[3]

Table S1: Hydrogen bond distances (Å) in structural complexes along the PES of HIV-1 PR—substrates systems (Figure 8) for the general acid-base mechanism.

H-Bond	INT	TS	Prod
O_{WatC}—$H1_{WatC}$--$OD1_{Asp25}$	2.30[4] 2.70[5]		2.65[6] 2.90[5]
OD1—H_{Asp25}'--N	2.50[4]		
$OD2_{Asp25}$--O_{WatC}	2.50[7] 2.91[8] 3.10[5]	2.56[9]	2.45[6] 2.80[5]
$OD1_{Asp25}$--$OD1_{Asp25}$'	2.60[9] 2.48[7] 2.80[5]	2.98[9]	2.30[10] 2.76[6]
$OD2_{Asp25}$'--N	2.80/2.90[7] 3.10[5]		2.90[10]
OD1--H_{Asp25}	1.13[9]	1.34[9]	
$HD1_{Asp25}$--$OD1_{Asp25}$'	1.69[9]	2.32[9]	
$HD1_{Asp25}$--$OD1_{Asp25}$	1.06[9]		
C_{SUB}--N_{SUB}	1.58[9]		2.58[6] 2.46[10]
C_{SUB}--O_{WatC}	1.49[9]		
O_{WatC}--$OD1_{Asp25}$'	3.00[5]		2.70[5]
O_{WatC}--$OD2_{Asp25}$'	3.00[5]		3.00[5]
N_{SUB}--$OD2_{Asp25}$'	2.35[9]		

N_{SUB}--$H2_{WatC}$	1.14[9]		
$H2_{WatC}$--$OD2_{Asp25}$'	1.18[9]	1.57[9]	
$H2_{WatC}$--O_{WatC}		1.30[9]	
CO_{SUB}--$OD2_{Asp25}$'	2.20[8]		
CO_{SUB}--$OD2_{Asp25}$	2.30[5]		2.51[10]
O_{WatC}--$OD2_{Asp25}$	2.91[8]		

Values are taken from both *in vitro* and *in silico* studies. ES = enzyme-substrate complex, INT represents tetrahedral intermediate, TS= transition state involved in the breakdown process of the INT and Prod= product.

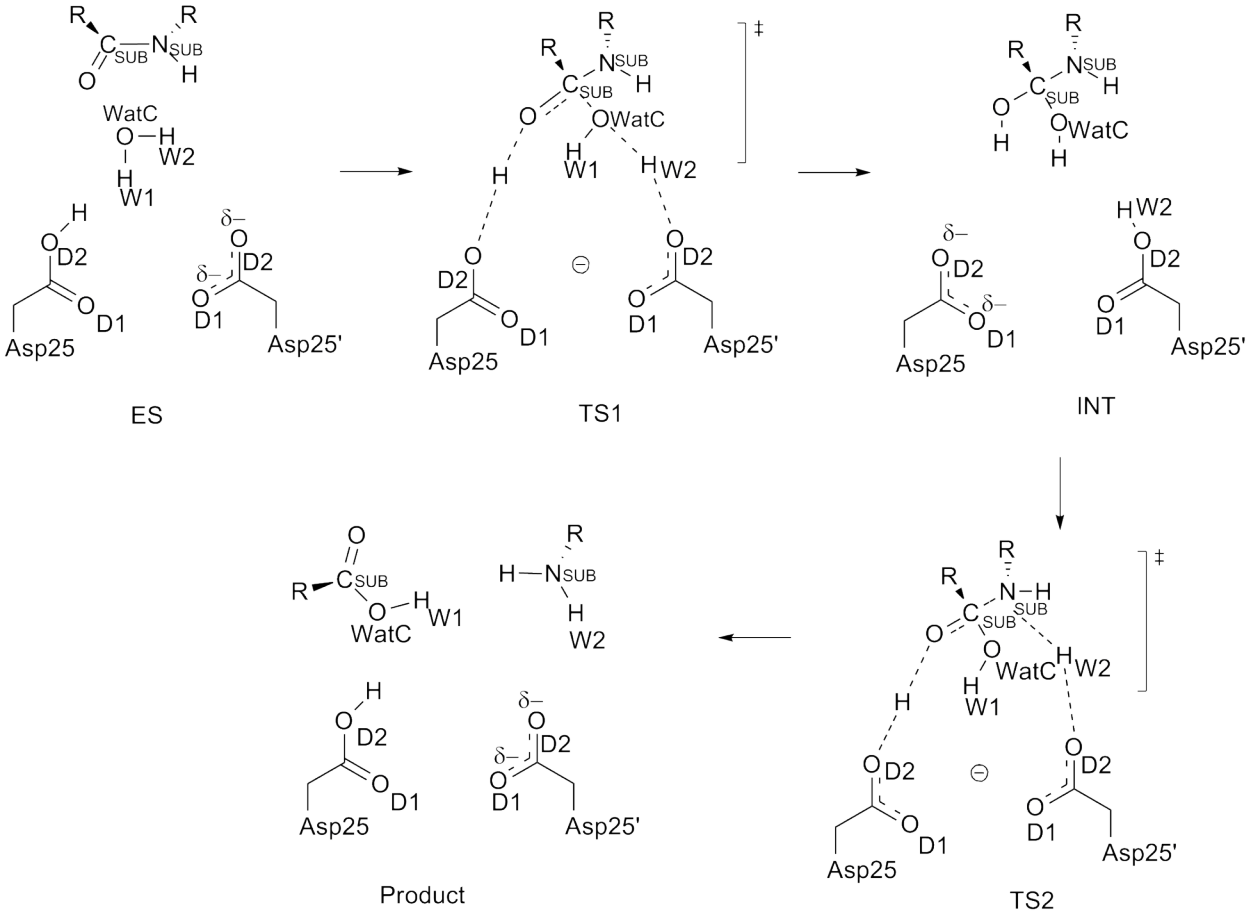


Figure S3: Complexes along mechanistic pathway of HIV-1 PR—substrates system.

Table S2: Theoretical and experimental interatomic distances (in Å) for important complexes (Figure S3) along the PES of HIV-1 PR—substrates. TS1 is obtained with ONIOM[11, 12] method at 6-31G(d)/AMBER[13] level of theory in this present study.

Distance	ES	TS1	INT	TS2	Prod
C_{SUB}—N_{SUB}	1.44[8]	1.38	1.70[4] 2.27[10]	1.52[14]	2.67[10] 1.65[7] 3.50[6] 2.02[14]
O_{WatC}—$HD2_{Asp25/25'}$	2.00[15]		≥1.50[4]		
C_{SUB}—O_{SUB}		1.34	1.80[4]		
C_{SUB}—O_{WatC}	2.91[15]	1.71	1.50[4]	1.88[4]	
$OD2_{Asp25'}$—O_{WatC}	2.80[15]	2.77			
$OD2_{Asp25}$—O_{WatC}	2.66[15]	2.97			
$OD1_{Asp25}$—O_{WatC}	3.28[15]	3.73			
$OD1_{Asp25'}$—O_{WatC}	2.96[15]	3.24			
N_{SUB}—$HD2_{Asp25'}$			2.00–3.00[4]	1.35[4] 1.26[14]	
OD2—$HD2_{Asp25'}$				1.32[14]	

ES is the enzyme—substrate complex, INT is the tetrahedral intermediate, TS1 and TS2 is the forming and breakdown process of the INT, respectively.

REFERENCES

[1] Krzeminska, A., Moliner, V., and Swiderek, K. (2016) Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease, *Journal of the American Chemical Society 138*, 16283-16298.

[2] Garrec, J., Sautet, P., and Fleurat-Lessard, P. (2011) Understanding the HIV-1 protease reactivity with DFT: what do we gain from recent functionals?, *The Journal of Physical Chemistry B 115*, 8545-8558.

[3] Park, H., Suh, J., and Lee, S. (2000) Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism, *Journal of the American Chemical Society 122*, 3901-3908.

[4] Piana, S., Bucher, D., Carloni, P., and Rothlisberger, U. (2004) Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations, *The Journal of Physical Chemistry B 108*, 11139-11149.

[5] Shen, C. H., Tie, Y. F., Yu, X. X., Wang, Y. F., Kovalevsky, A. Y., Harrison, R. W., and Weber, I. T. (2012) Capturing the Reaction Pathway in Near-Atomic-Resolution Crystal Structures of HIV-1 Protease, *Biochemistry 51*, 7726-7732.

[6] Bihani, S., Das, A., Prashar, V., Ferrer, J. L., and Hosur, M. V. (2009) X-ray structure of HIV-1 protease in situ product complex, *Proteins-Structure Function and Bioinformatics 74*, 594-602.

[7] Kumar, M., Prashar, V., Mahale, S., and Hosur, M. V. (2005) Observation of a tetrahedral reaction intermediate in the HIV-1 protease–substrate complex, *Biochem J 389*, 365-371.

[8] Das, A., Mahale, S., Prashar, V., Bihani, S., Ferrer, J. L., and Hosur, M. V. (2010) X-ray Snapshot of HIV-1 Protease in Action: Observation of Tetrahedral Intermediate and Short Ionic Hydrogen Bond SIHB with Catalytic Aspartate, *Journal of the American Chemical Society 132*, 6366-6373.

[9] Trylska, J., Grochowski, P., and McCammon, J. A. (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease, *Protein Science 13*, 513-528.

[10] Das, A., Prashar, V., Mahale, S., Serre, L., Ferrer, J. L., and Hosur, M. V. (2006) Crystal structure of HIV-1 protease in situ product complex and observation of a low-barrier hydrogen bond between catalytic aspartates, *P Natl Acad Sci USA 103*, 18464-18469.

[11] Svensson, M., Humbel, S., Froese, R. D., Matsubara, T., Sieber, S., and Morokuma, K. (1996) ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels− Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition, *The Journal of Physical Chemistry 100*, 19357-19363.

[12] Chung, L. W., Sameera, W., Ramozzi, R., Page, A. J., Hatanaka, M., Petrova, G. P., Harris, T. V., Li, X., Ke, Z., and Liu, F. (2015) The ONIOM method and its applications, *Chem Rev 115*, 5678-5796.

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

[14] Kipp, D. R., Hirschi, J. S., Wakata, A., Goldstein, H., and Schramm, V. L. (2012) Transition states of native and drug-resistant HIV-1 protease are the same, *P Natl Acad Sci USA 109*, 6543-6548.

[15] Okimoto, N., Tsukui, T., Kitayama, K., Hata, M., Hoshino, T., and Tsuda, M. (2000) Molecular Dynamics Study of HIV-1 Protease− Substrate Complex: Roles of the Water Molecules at the Loop Structures of the Active Site, *Journal of the American Chemical Society 122*, 5613-5622.

Structural Chemistry
https://doi.org/10.1007/s11224-018-1251-9

ORIGINAL RESEARCH

Unraveling the concerted catalytic mechanism of the human immunodeficiency virus type 1 (HIV-1) protease: a hybrid QM/MM study

Monsurat M. Lawal[1] · Zainab K. Sanusi[1] · Thavendran Govender[1] · Gideon F. Tolufashe[1] · Glenn E. M. Maguire[1,2] · Bahareh Honarparvar[1] · Hendrik G. Kruger[1]

Received: 5 October 2018 / Accepted: 26 November 2018

Abstract
We give an account of a one-step concerted catalytic mechanism of HIV-1 protease (PR) hydrolysis of its natural substrate using a hybrid QM/MM method. The mechanism is a general acid–base model having both catalytic aspartate groups participating and a water molecule attacking the natural substrate synchronously. Three different pathways were investigated: a concerted acyclic transition state (TS) mechanistic route, a concerted 6-membered cyclic TS process involving one water molecule, and another 6-membered ring TS pathway involving two water molecules. Activation free energies of approximately 15.2 and 16.6 kcal mol^{-1} were obtained for both concerted acyclic and the other possible reaction pathway involving two water molecules in the active site, respectively. The activation free energies are comparable to experimentally derived data of 15.69 kcal mol^{-1}. The outcome of the present work provides a plausible theoretical benchmark for the concerted enzymatic mechanism of HIV-1 PR and can be applied to related enzymatic processes.

Keywords HIV-1 protease · Natural substrate · QM/MM (Our own N-layered Integrated molecular Orbital and molecular Mechanics ONIOM) method · Concerted transition states · Catalytic mechanism

Introduction

One infectious disease that has had both a profound health and cultural impact on the human race in recent decades is the acquired immune deficiency syndrome (AIDS) [1] caused by the human immunodeficiency virus (HIV) [1, 2], and it remains one of the severest health challenges [2]. HIV attacks its host's immune system thus making such individuals vulnerable to any infection; a breakthrough in the treatment of HIV-1 is the use of drugs inhibiting specific enzymes necessary for the replication of the virus [3]. Among these enzymes is HIV-1 protease (PR), which is an important degrading enzyme necessary for the proteolytic cleavage of the Gag and Gag-Pol polyproteins, required for the development of mature virion proteins [4–7]. The hydrolytic action of the PR on these asymmetric natural polyprotein substrate sequences results in the processing of the corresponding mature virion [8, 9]. The mechanism of action of PR has been a subject of research over the past three decades [10].

The catalytic mechanism of the HIV-1 PR is one of the most studied aspartate protease reactions. Both experimental and theoretical techniques have been harnessed to provide a better understanding on a number of possible reaction pathways for the catalytic cleavage of the natural substrate/ligand by the PR [11–15]. The aspartate dyad of the HIV-1 PR is most often monoprotonated [16] in such theoretical studies (Fig. 1).

Most of the recent studies [14, 17–26] have investigated the stepwise general acid–base mechanism involving catalytic water (WatC) at the active site of the HIV-1 PR, whereby the hydrolysis occurs in several steps (Scheme 1). Another possible mechanistic pathway is a nucleophilic process [19, 20, 26], which is a variance of the stepwise general acid–base mechanism without WatC. Researchers have obtained both theoretical

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11224-018-1251-9) contains supplementary material, which is available to authorized users.

✉ Hendrik G. Kruger
kruger@ukzn.ac.za

[1] Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa

[2] School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa

Published online: 14 December 2018

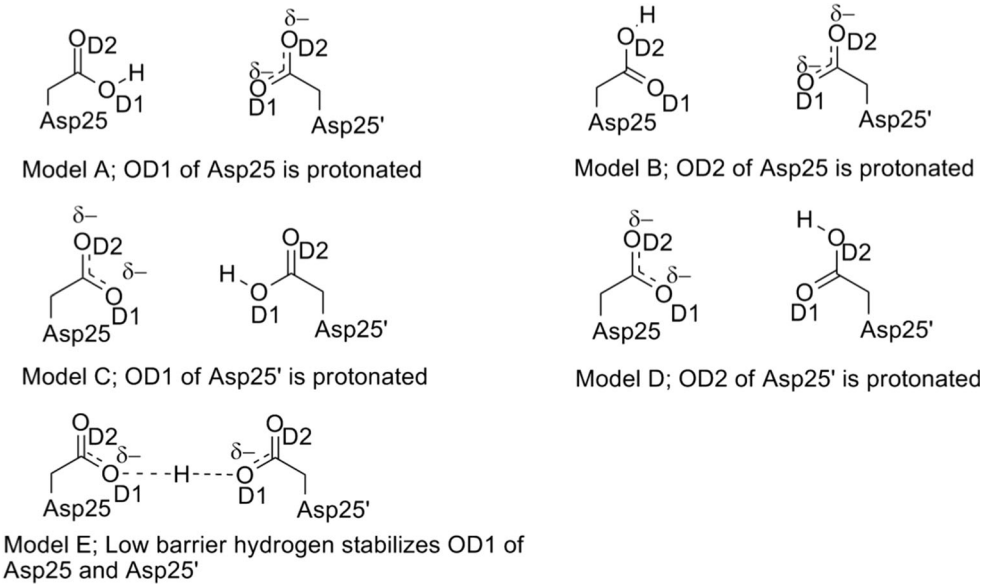
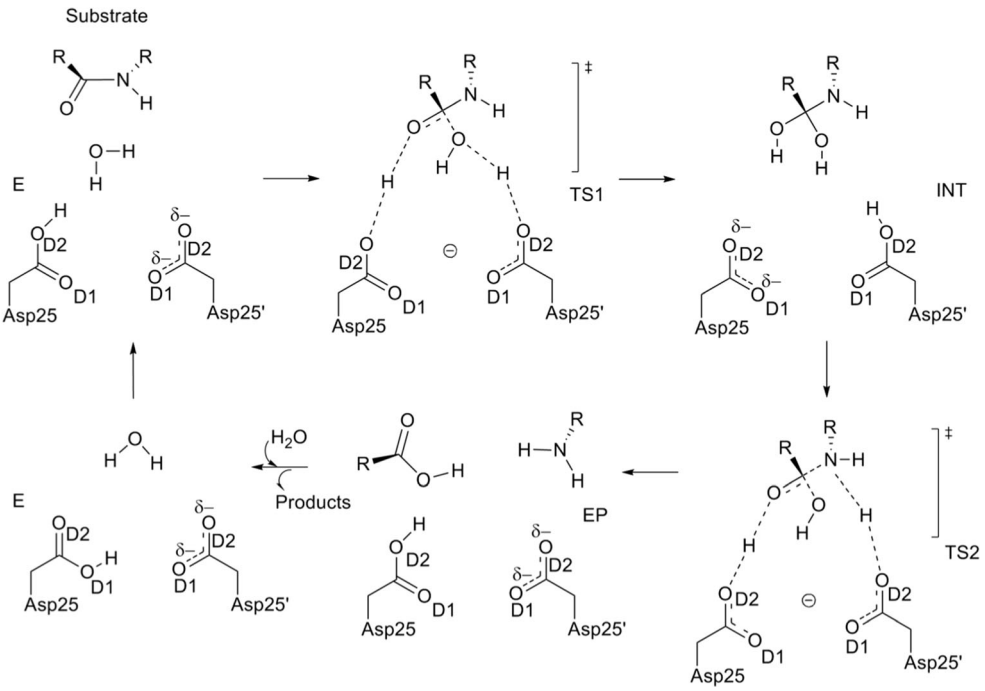


Fig. 1 Possible monoprotonated models for the catalytic aspartate dyad

and experimental rate-determining transition-state (TS) energies [14, 17–26] that correlate with experimental energies of peptide/substrate hydrolysis for both stepwise and nucleophilic mechanisms. Despite extensive investigations on HIV-1 PR catalytic pathways, little attention has been given to a synchronous model in which the proteolytic reaction could occur as a one-step concerted process.

In 1991, however, Jaskólski et al. [13] proposed a one-step concerted mechanism for the reaction; the nucleophile (water molecule) and electrophile (an acidic proton) attack the scissile bond in a concerted manner. At the beginning of the reaction, the acidic proton is located on the OD2 atom of that aspartate, which is proximal to the nitrogen atom of the scissile amide [13]. The post-reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms (Model E; Fig. 1). The proposed concerted mechanism in this study was indeed derived based on the reported experimental protocol [13] which is summarized in Scheme S1 of the supporting information (SI).

We found only two theoretical models [14, 19] in literature with respect to the one-step concerted perspective [13] having cyclic TSs with activation free energies ($\Delta G^{\ddagger}$) of 30 [19] and



Scheme 1 Stepwise general acid–base catalytic mechanism of HIV-1 PR [27]. ES is the enzyme–substrate complex. TS1 is the first transition structure which should be overcome before the formation of the tetrahedral intermediate (INT). TS2 is the breakdown of the INT complex while EP is the separated products. Adapted from literature [27]

43.5 [14] kcal mol^{-1}, respectively. These values are higher than experimental $\Delta G^{\ddagger}$ values (between 14.86 and 21.03 kcal·mol^{-1}) depending on the amino acid sequence [11, 12, 28–33]. In light of literature support for theoretical 6-membered ring transition structures [34–41], a one-step cyclic concerted chemical process (Scheme 2) appears plausible and can potentially provide a new understanding of the catalytic mechanism of HIV-1 PR.

Method

System setup

The X-ray crystal structure of MA-CA substrate (code: 1KJ4) [42] was obtained from RCSB PDB [43] and complexed with HIV-1 PR (code: 1A30) [44]. The catalytic water molecule was manually inserted at the active site and the bond distances observed from the 1LV1 structure were employed [45] (an apo HIV-1 PR co-crystallized with a water molecule at the active site). This complex was subjected to classical molecular dynamics (MD) simulation (20 ns long) as described in our previous work [46]. The lowest energy structure from the MD run was partitioned into two layers and Our Own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM) [47] (QM/MM) approach was applied to investigate the mechanism of the reaction. This was done after stripping off the explicit solvation box and non-required atoms (Cl–) inherent from the MD simulation. The catalytic active site, natural substrate (Matrix-Capsid segment (MA-CA)), and water were placed at a high layer (B3LYP [48, 49]/6-31++G(d,p) [50]) while the remaining residues were at the low layer (AMBER) [51] for geometry optimization (Fig. 2).

The reliability and accuracy of the B3LYP/6-31G(d):AMBER ONIOM model has been established for a similar study [52, 53]. In order to improve the level of accuracy of the results and consider the effect of diffusion and polarization functions on the hydrogen atoms involved at the reactive center, B3LYP/6-31++G(d,p):AMBER was used. Given the fact that one of the aspartate residues is protonated (Fig. 1) for a concerted mechanism where both the nucleophile (water molecule) and electrophile (an acidic proton) are expected to attack the scissile bond in a concerted manner, we protonate OD2 of Asp25′ (Model D; Fig. 1) as proposed from experiment [13].

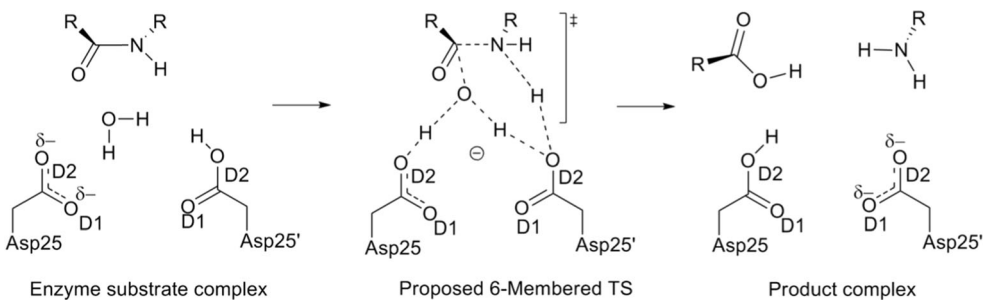
Transition state modeling and energy calculations

The complex obtained (lowest energy snapshot) from MD simulations was subjected to constraining and relaxing for the TS QM/MM modeling. Interatomic distances for the proposed cyclic transition state were constrained (Owat–Csub at 2.40 Å, Nsub–OD2 Asp25′ at 2.79 Å and Owat–OD2 Asp25′ at 2.71 Å, Scheme 2) during an optimization to find a suitable TS starting structure. All constraints were removed for the subsequent transition state optimization.

Calculations proceeded with a full geometry optimization of all structures using the selected two-layered ONIOM (B3LYP/6-31++G(d,p):AMBER) model. Vibrational frequencies [54] were computed for the various species to characterize them as local minima (no negative eigenvalues) and the TS structure having exactly one imaginary frequency. Intrinsic reaction coordinates (IRCs) [55] were computed to verify the transition structures are truly the lowest saddle points connecting the expected reactant and product complexes on the reaction pathway. GaussView 5.0.8 [56] was used as pre-processor and post-processor visual interface for this study, while all calculations were executed within the Gaussian 09 program package [57].

Results and discussion

The first attempt to find a cyclic TS (Scheme 2) gave us in fact a concerted acyclic TS involving one water molecule (Scheme 3). Refinement of the cyclic TS starting structures enabled us to find them as well. The relative thermodynamic and kinetic parameters for the concerted acyclic and cyclic mechanistic cleavage of MA-CA by HIV-1 PR are presented in Table 1. Generally, all the studied concerted mechanistic pathways are exergonic with highly negative free energies at product formation (Schemes 2, 3, and 4). The change in total energy (ΔE), enthalpy (ΔH), and entropy (TΔS) values of the ES and product complexes are negative. The free energy (ΔG) values related to the difference between ΔH and TΔS are also



Scheme 2 Proposed concerted 6-membered cyclic enzymatic mechanism of HIV-1 PR

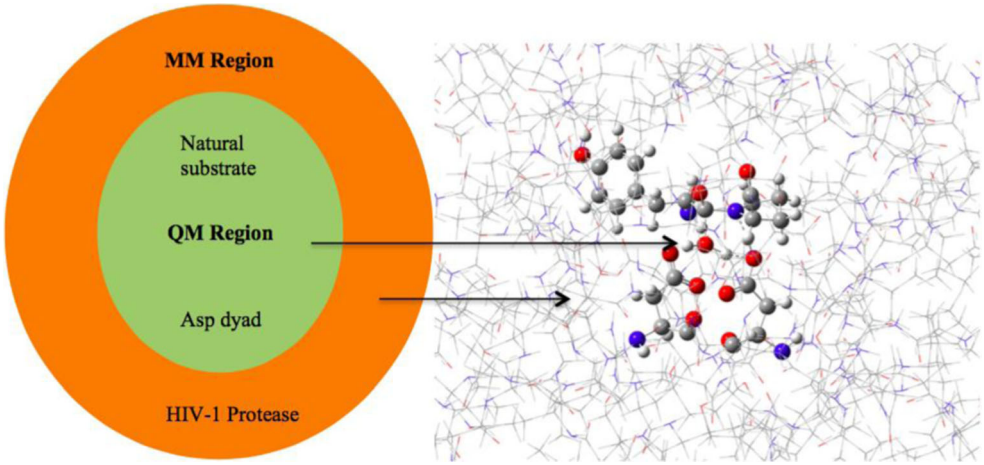


Fig. 2 Schematic representation of the two-layered ONIOM (DFT:AMBER) model of HIV-1 PR–MA-CA complex. ONIOM output files for the TS structures are available in PDB format with the SI

negative and ΔG of these minima seems driven by ΔH with characteristic large negative values ranging between − 33 and − 60 kcal mol^{-1}.

Cleavage of the substrate by HIV-1 PR gave a significantly large level of disorderliness as observed from the activation entropy ($T\Delta S^{‡}$) of the TS structures having more negative values than the activation enthalpy, $\Delta H^{‡}$ (Table 1). Activation free energies ($\Delta G^{‡}$) of the TS structures are thus favored by large entropy penalties of − 16.16, − 38.71, and − 54.92 kcal mol^{-1} for the 6-membered cyclic, acyclic, and 6-membered cyclic (with two water) models, respectively. The sum of the total free energy of each mechanistic pathway gave − 28.71, − 13.87, and − 46.48 kcal mol^{-1} for the 6-membered cyclic, acyclic, and 6-membered cyclic (with two water) systems, respectively.

Based on thermodynamics (more negative ΔH) and overall system energetics (ΔE), the two-water mechanistic pathway is the most favorable; addition of another water molecule at the HIV-1 PR active site essentially pulls down the activation barrier and improves the cleavage process. However, when the $\Delta G^{‡}$ of the TS structures were considered, the concerted acyclic mechanism involving a catalytic water molecule proved to be the most favored process (Table 1). This is a well-established perspective from literature using both in vitro and in silico methods [10]. The reaction rate of the TS structures depicted by lnk follows the order C-Ac-TS > C-6-TS > C-6-TS-TW.

Concerted acyclic general acid–base HIV-1 PR–substrate mechanistic pathway

The concerted acyclic TS (Scheme 3) leads to the expected reactant and product complexes (from the IRC calculation, Fig. 3) with an observed $\Delta G^{‡}$ of 15.21 kcal mol^{-1} (Table 1) at B3LYP/6-31++G(d,p):AMBER level of theory. During this concerted general acid–base mechanism, water donates one of its protons to the unprotonated Asp25 and the protonated Asp25′ loses its proton to the scissile nitrogen atom; thus, the initial acidic Asp becomes basic and vice versa. Meanwhile, the nucleophilic water (OH) attacks the scissile carbon resulting in substrate cleavage (Scheme 3).

The reaction involves bond forming and breaking processes; the initial C–N bond increases from 1.47 to 1.67 Å, at the TS and becomes 2.65 Å at the product complex, which is comparable to experimental distance of 2.70 Å between these atoms [58]. The proton transfer (to Asp25 OD2) from the catalytic water increases in bond length from 0.96 to 1.26 Å in the TS structure, thereby, moving closely to the OD2 of Asp25 with a bond distance of 1.14 Å. The scissile nitrogen atom accepts a proton from the protonated aspartate (Asp25′) with a bond distance of 1.22 Å (Fig. 4).

For the acyclic mechanism, the ES gave ΔE, ΔH, and $T\Delta S$ values of approximately − 40, − 39, and − 22 kcal mol^{-1}, respectively, with ΔG formation of − 17 kcal mol^{-1} (Table 1). This pre-ordered ES complex

Scheme 3 Concerted acyclic enzymatic mechanism of HIV-1 PR with its substrate

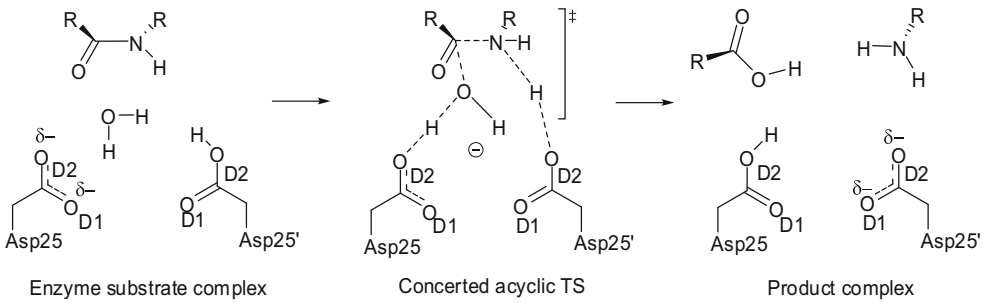


Table 1 Relative thermodynamic and kinetic parameters for the one-step catalytic mechanism of natural substrate (MA-CA) using ONIOM (B3LYP/6-31++G(d,p):AMBER)

	ΔE	ΔH	TΔS	$\Delta G^{\ddagger}$	lnk
Concerted acyclic mechanism					
ES	−40.28	−38.88	−21.66	−17.22	
C-Ac-TS	−28.60	−23.49	−38.71	15.21	3.77
PC	−34.71	−33.44	−10.74	−26.70	
Concerted 6-membered cyclic mechanism − 13.87					
C-6-TS	9.45	10.23	−16.16	26.39	−15.11
PC	−43.15	−41.75	−18.71	−23.04	
Concerted 6-membered cyclic mechanism with two water molecules					
ES-TW	−62.51	−59.66	−22.12	−37.54	
C-6-TS-TW	−35.82	−38.35	−54.92	16.57	1.49
PC-TW	−54.63	−53.37	−27.86	−25.51	

Values are reported in kilocalories per mole relative to the sum of separated reactants. *ES* enzyme–substrate complex, *TS* transition state structure, *PC* product complex, *TW* two water, *C* concerted, *Ac* acyclic. ONIOM output files for the TS structures are available in PDB format with the SI

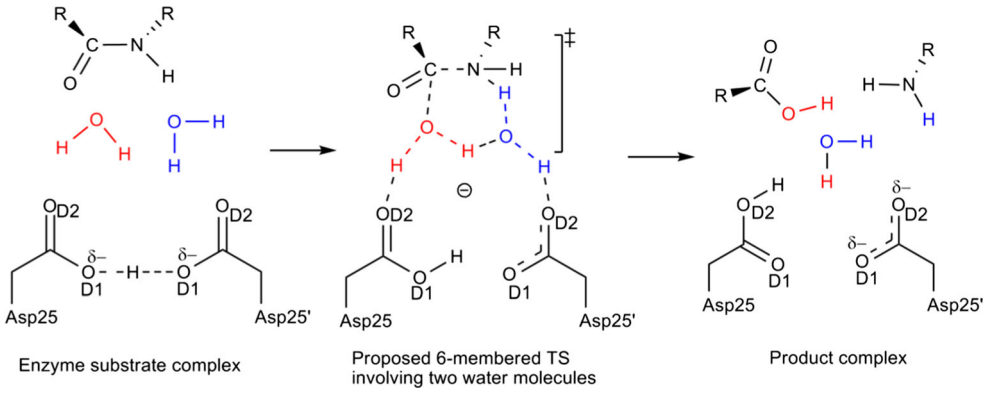
facilitates the concerted nature of the acyclic TS (Fig. 4) with an exergonic total $\Delta E^{\ddagger}$ of − 28.6 kcal mol^{-1}, $\Delta H^{\ddagger}$ value of − 23.5 kcal mol^{-1}, $T\Delta S^{\ddagger}$ of − 38.7 kcal mol^{-1}, and an activation barrier ($\Delta G^{\ddagger}$) of 15.2 kcal mol^{-1}. C-Ac-TS is more favored with entropy compared to enthalpy while the opposite holds for its PC. The calculated $\Delta G^{\ddagger}$ of 15.21 kcal mol^{-1} is in good agreement with experimentally deduced $\Delta G^{\ddagger}$ of 15.69 kcal mol^{-1} [30]. The overall mechanism remained exergonic with the product complex having a ΔG of − 26.70 kcal mol^{-1} (Fig. 3 and Table 1) with a total ΔG value of − 28.71 kcal mol^{-1} for the overall mechanistic process.

The concerted cyclic general acid–base HIV-1 PR–substrate mechanistic pathway

Subsequently, we also found the concerted 6-membered ring TS structure facilitated by one water molecule (Scheme 2), as well as another cyclic TS involving two water molecules (Scheme 4). Although the mechanism closely mimics the acyclic model, a slight difference was observed. Unlike the acyclic pathway in which the aspartates participate in bond sharing with the water and substrate, the 6-membered cyclic mechanism (involving one and two water molecules) showcased distinct characteristics.

In the case of one water cyclic mechanism (Scheme 2), the unprotonated Asp25 only acts as an anchoring entity rather than a base and a product complex is formed with Asp25′ protonated from water. In other words, only Asp25′ partakes in bond sharing and acts as an acid (donating its proton to the scissile nitrogen) as well as a base (protonated by water) at the end of the cleavage process. The mechanism, therefore, starts and ends with protonation model D (Fig. 1). C–N scissile bond of the C-6-TS was 1.65 Å and increased to 2.66 Å after substrate cleavage (PC). The interatomic distance between the scissile carbonyl (C=O) and the nucleophilic water (OH) was 1.75 Å while the second H of water was 1.43 Å away from the OD2 of Asp25 for the transition state (Fig. S1). The scissile nitrogen and proton from the Asp25′ gave an interatomic distance of 1.32 Å in the TS structure.

Optimization of the complexes down the reaction profile also yielded the ES and product complexes in the 6-membered one water cyclic mechanism with the same ES as the acyclic model (Table 1). The calculated thermodynamic parameters for the C-6-TS were slightly endergonic with $\Delta E^{\ddagger}$ of 9.45 kcal mol^{-1}, $\Delta H^{\ddagger}$ value of 10.23 kcal mol^{-1}, $T\Delta S^{\ddagger}$ of − 16.16 kcal mol^{-1}, and $\Delta G^{\ddagger}$ of 26.39 kcal mol^{-1}. This TS structure is largely driven by entropy contributions and the overall mechanistic pathway remained exergonic with a summed ΔG value of − 13.87 kcal mol^{-1} and the calculated ΔG value for PC is − 23.04 kcal mol^{-1}. Although, the activation free energy value is not perfectly comparable with experiment, it is worth mentioning that this 6-membered cyclic model is an improved model when compared to the previous contributions from literature [14, 19]. Krzemińska et al. [14] recently explored a 4-membered ring TS model devoid of the catalytic (or general acid/base) function of the aspartic moieties with a much higher free energy barrier of 43.5 kcal mol^{-1}.



Scheme 4 Proposed reaction scheme for the two-water mediated cleavage of natural substrate by HIV-1 PR

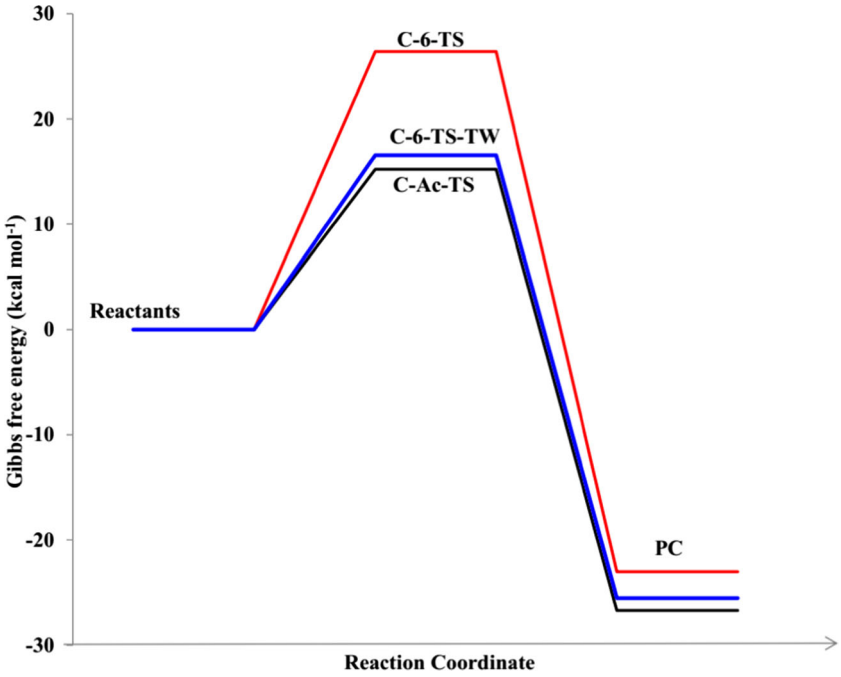


Fig. 3 Free energy profile for the one-step concerted cyclic and acyclic catalytic mechanism of HIV-1 PR and MA-CA natural substrate using ONIOM (B3LYP/6-31++G(d,p):AMBER) method. TS, transition state structure; TW, two water; PC, product complex; C, concerted; Ac, acyclic. ONIOM output files for the TS structures are available in PDB format with the SI

In order to improve the 6-membered cyclic model with catalytic water, we proposed a concerted 6-membered ring mechanistic pathway involving two water molecules (Scheme 4). The mechanism involves both aspartates acting as anchoring moieties (they are not involved in bond sharing with the substrate); however, their catalytic effect was still obvious with $\Delta G^{\ddagger}$ of 16.57 kcal mol^{-1} for the TS structure (Fig. 3 and Table 1). This calculated $\Delta G^{\ddagger}$ is in reasonable agreement with experimentally derived $\Delta G^{\ddagger}$ 15.69 kcal mol^{-1} for MA-CA (Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln) hydrolysis by HIV-1 PR [30]. This mechanism does not only showcase the success of a cyclic synchronous model but also offer another perspective on the importance of water molecules at the active site of the HIV-1 PR [59–62] in which the activation barrier was lowered by almost 10 kcal mol^{-1} in comparison to the one water model (Fig. 3).

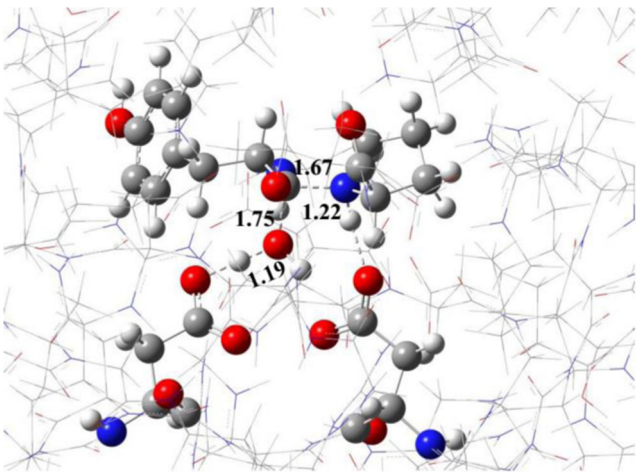


Fig. 4 The concerted acyclic TS of HIV-1 PR–MA-CA general acid–base mechanism using ONIOM (DFT:AMBER) method. Picture generated with GaussView 5.0.8 [56], distances are in angstrom (Å). ONIOM output files for the TS structures are available in PDB format with the SI (Figs. S1 and S2)

The calculated values for ΔE, ΔH, and $T\Delta S$ were − 62.51, − 51.66, and − 22.12 kcal mol^{-1}, respectively; ES complex of the two-water model with an estimated ΔG value of − 37.54 kcal mol^{-1} (Table 1). C-6-TS-TW has the highest $T\Delta S^{\ddagger}$ value of − 54.94 kcal mol^{-1} in comparison to the two other TS possibilities; this is expected due to the increased atomic constituent of this TS and the induced disorderliness from an additional water molecule. This two-water mediated mechanistic pathway also gave the highest total ΔG value of − 46.48 kcal mol^{-1} for its entire process, thus establishing the feasibility of the mechanism theoretically. The calculated ΔG for its PC is − 25.51 kcal mol^{-1} while − 53.37 and − 27.86 kcal mol^{-1} were obtained for ΔH and $T\Delta S$, respectively, (Table 1).

The one-step concerted acyclic TS model provides a plausible theoretical model for the enzymatic mechanism of HIV-1 PR. Unlike previous computational attempts [14, 19] for this mechanism with higher free energy barriers for peptide hydrolysis, much lower energy barriers (15.21 and 16.57 kcal mol^{-1}) are obtained herein. This favorable energy could be attributed to the studied concerted TS models and the HIV-1 PR preference for large hydrophobic side chains at the P1 position of the natural target [63]; MA-CA scissile bond is located between Tyr-Pro (Fig. 2).

Conclusion

In conclusion, we have reported a new (theoretical) perspective on three possible concerted general acid–base mechanisms for the HIV-1 PR catalysis of its natural substrate. The mechanism that provides marginally the lowest activation barrier involves an acyclic TS model with one water molecule (Scheme 3) at the HIV-1 PR active site. We also proposed a two-water model (Scheme 4) involving cyclic TS structure having an observed activation free energy that is comparable to experiment and should be pursued experimentally in subsequent research. This present investigation could potentially provide a better understanding on achieving a single rate-limiting step for HIV-1 PR catalysis since the possibility of the existence of more than one rate-determining step has been proposed in the stepwise mechanistic pathway [64].

These models also provide new structural information about the exact nature of the protonation state of the two catalytic Asp residues, during hydrolysis of the natural substrate for a concerted general acid–base mechanism. The outcome of this study is quite informative and the TS model will be applied to related homodimeric protease and perhaps other enzymatic processes. Future studies will attempt to obtain a better understanding of the recognition phenomena of the HIV-1 PR towards natural substrates with preference for the scissile amide bonds.

Acknowledgements We are grateful to the Centre for High Performance Computing (www.chpc.ac.za) for computational resources.

Funding information The authors thank the College of Health Sciences, University of KwaZulu-Natal, Asphen Pharmacare, Medical Research Council, and the National Research Foundation (all in South Africa) for the financial support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Schuman JS, Orellana J, Friedman AH, Teich SA (1987) Acquired immunodeficiency syndrome (AIDS). Surv Ophthalmol 31:384–410
2. UNAIDS, https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics, (2017)
3. Potempa M, Lee S-K, Wolfenden R, Swanstrom R (2015) The triple threat of HIV-1 protease inhibitors. The future of HIV-1 therapeutics, springer, pp 203–241
4. Crawford S, Goff SP (1985) A deletion mutation in the 5′part of the pol gene of Moloney murine leukemia virus blocks proteolytic processing of the gag and pol polyproteins. J Virol 53:899–907
5. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon R, Scolnick EM, Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci 85:4686–4690
6. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature 331:280–283
7. Ashorn P, McQuade TJ, Thaisrivongs S, Tomasselli AG, Tarpley WG, Moss B (1990) An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. Proc Natl Acad Sci 87:7472–7476
8. Debouck C, Gorniak JG, Strickler JE, Meek TD, Metcalf BW, Rosenberg M (1987) Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. Proc Natl Acad Sci 84:8903–8906
9. Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu C-T, Lumma PK, Freidinger RM, Veber DF, Sigal IS (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. Biochem Biophys Res Commun 156:297–303
10. Lawal MM, Sanusi ZK, Govender T, Maguire GEM, Honarparvar B, Kruger HG (2018) From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets. Curr Med Chem
11. Hyland LJ, Tomaszek Jr TA, Roberts GD, Carr SA, Magaard VW, Bryan HL, Fakhoury SA, Moore ML, Minnich MD, Culp J (1991) Human immunodeficiency virus-1 protease. 1. Initial velocity studies and kinetic characterization of reaction intermediates by 18O isotope exchange. Biochemistry 30:8441–8453
12. Hyland LJ, Tomaszek Jr TA, Meek TD (1991) Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. Biochemistry 30:8454–8463
13. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SB, Wlodawer A (1991) Structure at 2.5-. ANG. Resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. Biochemistry 30:1600–1609
14. Krzeminska A, Moliner V, Swiderek K (2016) Dynamic and electrostatic effects on the reaction catalyzed by HIV-1 protease. J Am Chem Soc 138:16283–16298
15. Ribeiro AJM, Santos-Martins D, Russo N, Ramos MJ, Fernandes PA (2015) Enzymatic flexibility and reaction rate: a QM/MM study of HIV-1 protease. ACS Catal 5:5617–5626
16. Mcgee TD, Edwards J, Roitberg AE (2014) pH-REMD simulations indicate that the catalytic aspartates of HIV-1 protease exist primarily in a monoprotonated state. J Phys Chem B 118:12577–12585
17. Rodriguez EJ, Angeles TS, Meek TD (1993) Use of N-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus-1 protease. Biochemistry 32:12380–12385
18. Silva AM, Cachau RE, Sham HL, Erickson JW (1996) Inhibition and catalytic mechanism of HIV-1 aspartic protease. J Mol Biol 255:321–340
19. Park H, Suh J, Lee S (2000) Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism. J Am Chem Soc 122:3901–3908
20. Chatfield DC, Brooks BR (1995) HIV-1 protease cleavage mechanism elucidated with molecular-dynamics simulation. J Am Chem Soc 117:5561–5572
21. Lee H, Darden TA, Pedersen LG (1996) An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease. J Am Chem Soc 118:3946–3950

22. Liu HY, MullerPlathe F, vanGunsteren WF (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. J Mol Biol 261:454–469
23. Okimoto N, Makiyama M, Hata M, Tsuda M (1997) Gag protein hydrolysis mechanism by HIV-1 protease - investigation by semi-empirical molecular orbital method. J Chem Soc Jpn:260–266
24. Okimoto N, Tsukui T, Hata M, Hoshino T, Tsuda M (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: investigation by the ab initio molecular orbital method. J Am Chem Soc 121:7349–7354
25. Trylska J, Bala P, Geller M, Grochowski P (2002) Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. Biophys J 83:794–807
26. Trylska J, Grochowski P, McCammon JA (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. Protein Sci 13:513–528
27. Piana S, Carloni P, Parrinello M (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. J Mol Biol 319:567–583
28. Feher A, Weber IT, Bagossi P, Boross P, Mahalingam B, Louis JM, Copeland TD, Torshin IY, Harrison RW, Tozser J (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. Eur J Biochem 269:4114–4120
29. Tozser J, Bagossi P, Weber IT, Copeland TD, Oroszlan S (1996) Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases. J Biol Chem 271:6781–6788
30. Maschera B, Darby G, Palu G, Wright LL, Tisdale M, Myers R, Blair ED, Furfine ES (1996) Human immunodeficiency virus - mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. J Biol Chem 271:33231–33235
31. Kipp DR, Silva RG, Schramm VL (2011) Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease. J Am Chem Soc 133:19358–19361
32. Boross P, Bagossi P, Copeland TD, Oroszlan S, Louis JM, Tozser J (1999) Effect of substrate residues on the P2 ′ preference of retroviral proteinases. Eur J Biochem 264:921–929
33. Rodriguez EJ, Debouck C, Deckman IC, Abu-Soud H, Raushel FM, Meek TD (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry. Biochemistry 32:3557–3563
34. Kruger HG (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides. J Mol Struct-Theochem 577: 281–285
35. Gokul V, Kruger HG, Govender T, Fourie L, Power TD (2004) An ab initio mechanistic understanding of the regioselective acetylation of 8,11-dihydroxy-pentacyclo[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam. J Mol Struct-Theochem 672:119–125
36. Kruger HG, Mdluli P, Power TD, Raasch T, Singh A (2006) Experimental and computational studies of the regioselective protection of hydantoins using anhydride. J Mol Struct-Theochem 771: 165–170
37. Makatini MM, Petzold K, Sriharsha SN, Ndlovu N, Soliman MES, Honarparvar B, Parboosing R, Naidoo A, Arvidsson PI, Sayed Y, Govender P, Maguire GEM, Kruger HG, Govender T (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. Eur J Med Chem 46:3976–3985
38. Singh T, Kruger HG, Bisetty K, Power TD (2012) Theoretical study on the formation of a pentacyclo-undecane cage lactam. Comput Theor Chem 986:63–70
39. Rauf SMA, Arvidsson PI, Albericio F, Govender T, Maguire GEM, Kruger HG, Honarparvar B (2015) The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study. Org Biomol Chem 13:9993–10006
40. Lawal MM, Govender T, Maguire GE, Honarparvar B, Kruger HG (2016) Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study. J Mol Model 22:235
41. Fakhar Z, Govender T, Lamichhane G, Maguire GE, Kruger HG, Honarparvar B (2017) Computational model for the acylation step of the β-lactam ring: potential application for l, d-transpeptidase 2 in mycobacterium tuberculosis. J Mol Struct 1128:94–102
42. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, Structure 10:369–381
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
44. Louis JM, Dyda F, Nashed NT, Kimmel AR, Davies DR (1998) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. Biochemistry 37:2105–2110
45. Kumar M, Kannan K, Hosur M, Bhavesh NS, Chatterjee A, Mittal R, Hosur R (2002) Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. Biochem Biophys Res Commun 294:395–401
46. Fakhar Z, Govender T, Maguire GE, Lamichhane G, Walker RC, Kruger HG, Honarparvar B (2017) Differential flap dynamics in L, D-transpeptidase2 from Mycobacterium tuberculosis revealed by molecular dynamics. Mol BioSyst 13:1223–1234
47. Svensson M, Humbel S, Froese RD, Matsubara T, Sieber S, Morokuma K (1996) ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels− Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition. J Phys Chem 100:19357–19363
48. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37:785
49. Becke AD (1993) A new mixing of Hartree–Fock and local density-functional theories. J Chem Phys 98:1372–1377
50. Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA (2001) 6-31G* basis set for third-row atoms. J Comput Chem 22: 976–984
51. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
52. Sanusi ZK, Govender T, Maguire GE, Maseko SB, Lin J, Kruger HG, Honarparvar B (2018) An insight to the molecular interactions of the FDA approved HIV PR drugs against L38L↑ N↑ L PR mutant. J Comput Aided Mol Des 32:459–471
53. Sanusi Z, Govender T, Maguire G, Maseko S, Lin J, Kruger H, Honarparvar B (2017) Investigation of the binding free energies of FDA approved drugs against subtype B and C-SA HIV PR: ONIOM approach. J Mol Graph Model 76:77–85
54. Ochterski JW (1999) Vibrational analysis in Gaussian, help@gaussian. com
55. Gonzalez C, Schlegel HB (1990) Reaction path following in mass-weighted internal coordinates. J Phys Chem 94:5523–5527
56. Dennington R, Keith T, Millam J (2009) In GaussView, in, Semichem Inc., Shawnee Mission KS
57. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts

R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, in, Gaussian, Inc., Wallingford, CT. USA

58. Bihani S, Das A, Prashar V, Ferrer JL, Hosur MV (2009) X-ray structure of HIV-1 protease in situ product complex. Proteins: Struct, Funct, Bioinf 74:594–602
59. Liu Z, Wang Y, Yedidi RS, Dewdney TG, Reiter SJ, Brunzelle JS, Kovari IA, Kovari LC (2013) Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes. Biochem Biophys Res Commun 430:1022–1027
60. Li Z, Lazaridis T (2003) Thermodynamic contributions of the ordered water molecule in HIV-1 protease. J Am Chem Soc 125: 6636–6637
61. Wang Y-X, Freedberg DI, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Bhat TN, Erickson JW, Torchia DA (1996) Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. J Am Chem Soc 118: 12287–12290
62. Okimoto N, Tsukui T, Kitayama K, Hata M, Hoshino T, Tsuda M (2000) Molecular dynamics study of HIV-1 protease− substrate complex: roles of the water molecules at the loop structures of the active site. J Am Chem Soc 122:5613–5622
63. Tozser J, Zahuczky G, Bagossi P, Louis JM, Copeland TD, Oroszlan S, Harrison RW, Weber IT (2000) Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur J Biochem 267:6287–6295
64. Shen CH, Tie YF, Yu XX, Wang YF, Kovalevsky AY, Harrison RW, Weber IT (2012) Capturing the reaction pathway in near-atomic-resolution crystal structures of HIV-1 protease. Biochemistry 51: 7726–7732

Unravelling the concerted catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease: a hybrid QM/MM study

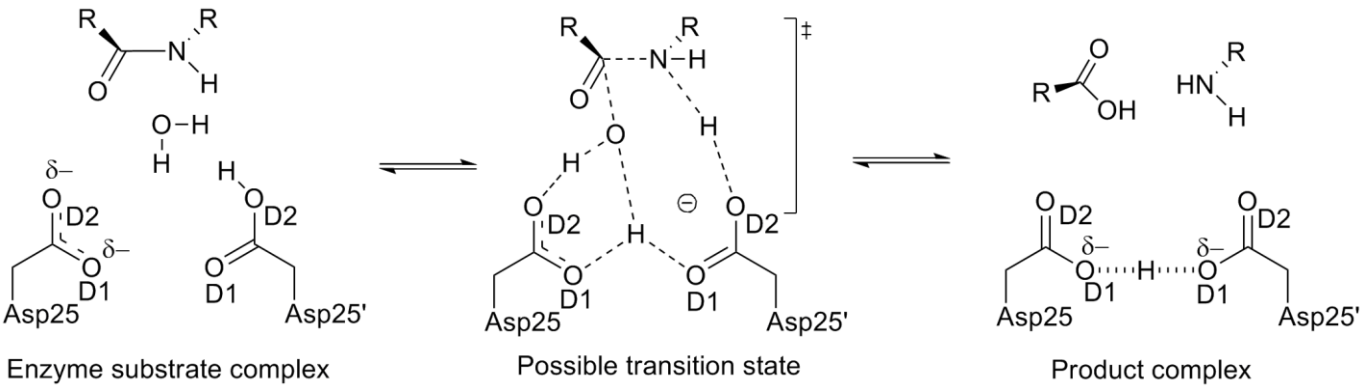
Supporting information

Monsurat M. Lawal,[a] Zainab K. Sanusi,[a] Thavendran Govender,[a] Gideon F. Tolufashe,[a] Glenn E. M. Maguire,[a, b] Bahareh Honarparvar,[a] and Hendrik G. Kruger.[a,*]

[a]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

[b]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.

*__Corresponding author__: Prof. H. G. Kruger; kruger@ukzn.ac.za, Fax: +27312601845, Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.



Scheme S1: Proposed concerted HIV-1 PR mechanism based on Jaskólski *et al.*[1] hypothesis.

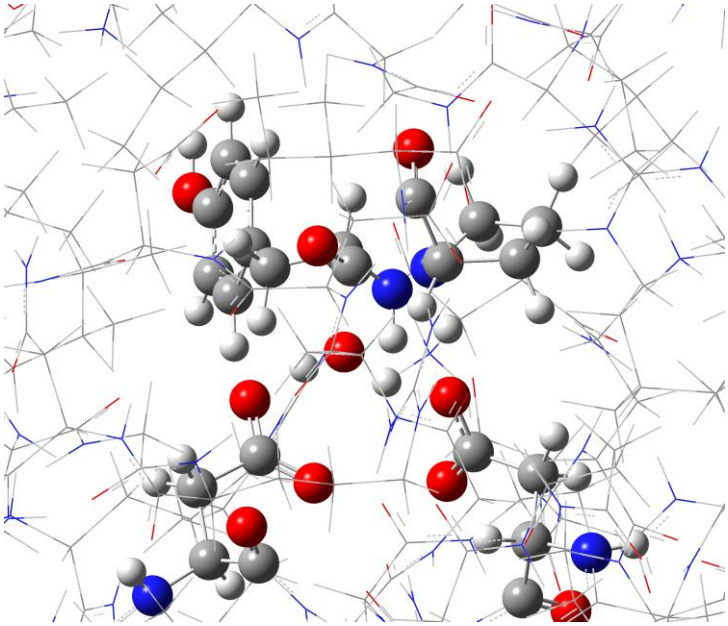


Figure S1: Concerted 6-membered cyclic TS involving one water molecule at the active site of HIV-1 PR—MA-CA complex

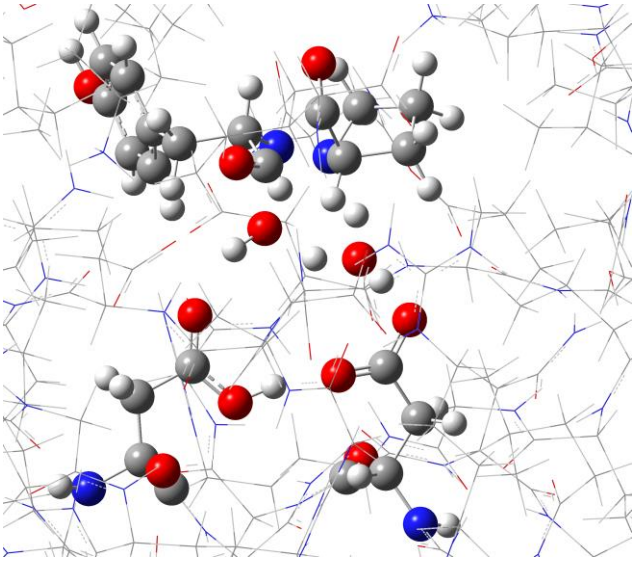


Figure S2: Concerted 6-membered cyclic TS involving two water molecules at the active site of HIV-1 PR—MA-CA complex

The PDB format of the ONIOM output files for the TS structures.

TS_PDB.zip

References

1. Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B.; Wlodawer, A., Structure at 2.5-. Ang. Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-Based Inhibitor. *Biochemistry* **1991**, *30*, 1600-1609.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 6389−6400

pubs.acs.org/JPCB

Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates

Zainab K. Sanusi,[†] Monsurat M. Lawal,[†] Thavendran Govender,[§] Glenn E. M. Maguire,[†,‡] Bahareh Honarparvar,[†] and Hendrik G. Kruger*[,†]

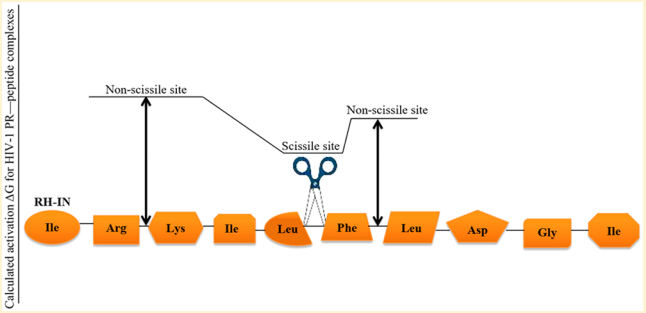
[†]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa

[§]Ansynth PTY LTD, 498 Grove End Drive, South Africa

[‡]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa

Supporting Information

ABSTRACT: The Human Immunodeficiency Virus type 1 (HIV-1) protease is a crucial target for HIV/AIDS treatment, and understanding its catalytic mechanism is the basis on which HIV-1 enzyme inhibitors are developed. Several experimental studies have indicated that HIV-1 protease facilitates the cleavage of the Gag and Gag-Pol polyproteins and it is highly selective with regard to the cleaved amino acid precursors and physical parameters. However, the main theoretical principles of substrate specificity and recognition remain poorly understood theoretically. By means of a one-step concerted transition state modeling, the recognition of natural substrates by HIV-1 PR subtypes (B and C-SA) was studied. This was carried out to compare the activation free energies at varying peptide bond regions (scissile and nonscissile) within the polypeptide sequence using ONIOM calculations. We studied both P3−P3′ and P5−P5′ natural substrate systems. For P3−P3′ substrates, excellent recognition was observed for the MA-CA family but not for the RH-IN substrates. Satisfactory recognition for the latter was only observed for the longer sequence (P5−P5′) after the substrate was subjected to an MD run to maximize the interaction between the enzyme and the substrate. These results indicate that both sequence and structure are important for correct scissile bond recognition of these natural substrates.



1. INTRODUCTION

A group of proteolytic enzymes with two aspartyl (Asp) residues at their active sites are known as the aspartic proteinases.[1,2] This family includes renin,[3,4] pepsin,[3,4] penicillopepsin,[5] and the human immunodeficiency virus protease (HIV-1 PR).[5,6] HIV-1 PR catalyzes the hydrolysis of Gag and Gag-Pol polyprotein precursors at specific sites (scissile bonds). These proteins fulfill essential roles in the maturation of the infectious virus that causes AIDS in humans.[7−9] The active form of HIV PR is a homodimer that normally contains 99 amino acids per monomer with a conserved catalytic triad (Asp25-Thr-26-Gly27) at the active site.[10−12] In the active site, the Asp25 of both monomers are directly involved in the catalysis and can exist in a number of protonated states.[13−17]

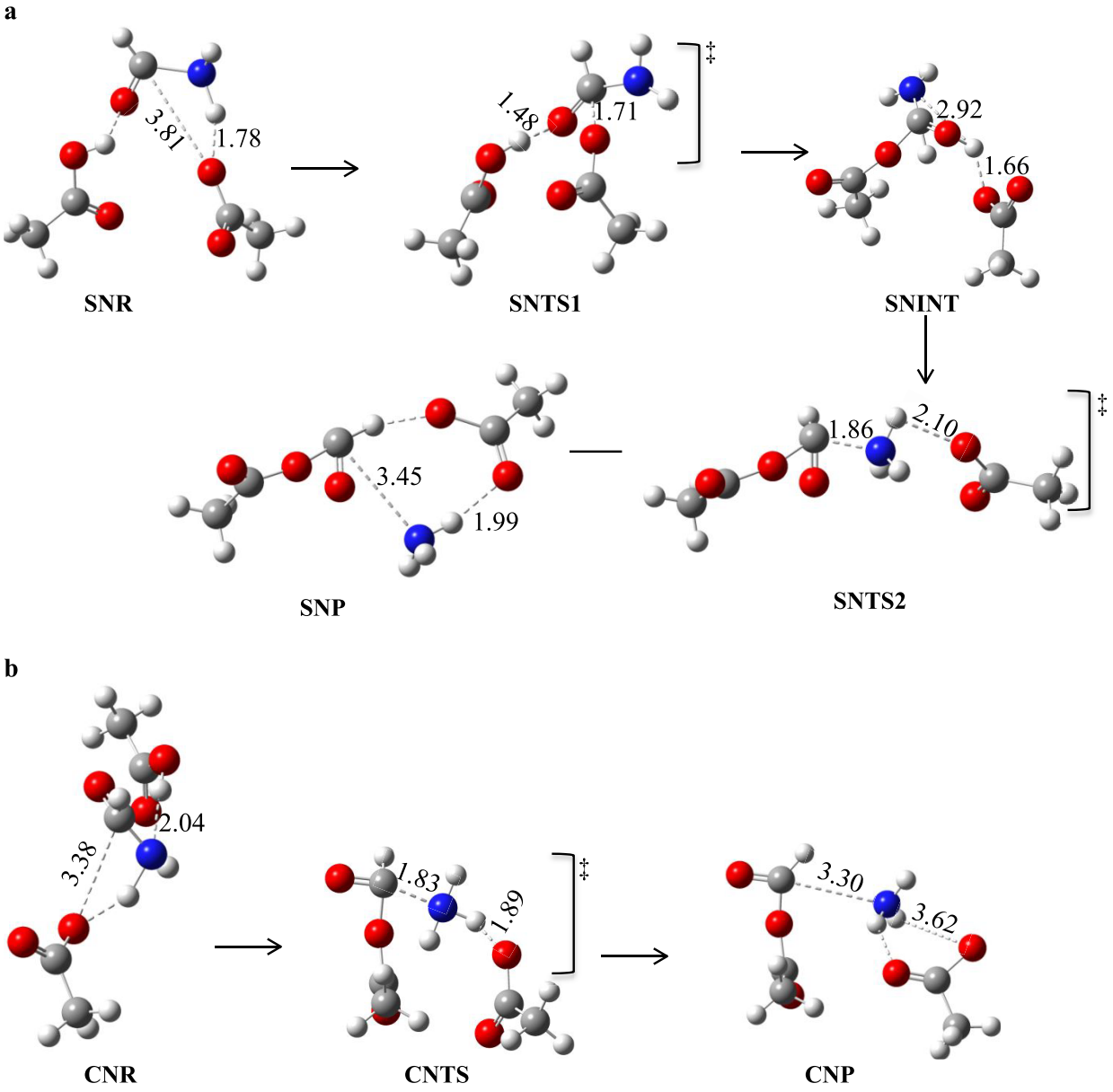
After the three-dimensional structure of the HIV-1 PR was resolved using single crystal X-ray analysis,[18,19] numerous studies were performed to design specific and potent inhibitors.[20−22] Despite that, the detailed mechanism of HIV-1 PR catalysis of the natural substrate is still a matter of controversy.[17] One factor that has impaired a comprehensive analysis is the different sequence homology of the cleavage sites on the natural substrates, making the substrate specificity determinants difficult to identify.[23] In addition, the cleavage site sequences that occur have been compared for HIV-1 versus HIV-2 and results reveal that both enzymes do not have a precise consensus substrate sequence.[24,25] Since several transition state analogues of the enzyme-catalyzed reaction were found to act as excellent inhibitors, it is crucial to fully understand the mechanism of the HIV-1 enzyme,[26] so that the design of new and improved analogues can be attempted.

Several HIV-1 PR cleavage mechanisms based on both *in vitro* and *in silico* research have been suggested.[5,6,15,20,27−29] However, there is no consensus on the most feasible mechanistic pathway. Most of these studies classify the mechanism into two general groups: the nucleophilic[6,12,29] and the general acid−general base processes.[20,30,31] The latter is the predominant mechanism from the literature for aspartic proteases and HIV-PR in particular.[17,28,32−35] The nucleophilic mechanism (Scheme 1) involves a direct attack on the scissile peptide (C−N) bond by one of the protonated catalytic aspartates that acts as the nucleophile[9] yielding an anhydride intermediate (stepwise nucleophilic reaction) or a transition

Received: March 8, 2019
Revised: June 30, 2019
Published: July 8, 2019

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

Scheme 1. Proposed (a) Stepwise and (b) Concerted Nucleophilic Reaction Mechanism for the HIV-1 PR–Substrate System (Redrawn from the Literature[6]) Involving a Simplified Peptide and Active Site Model at the MP2/6-31G(2d)//RHF/6-31G(2d) Level of Theory[a]



[a]The legends are SNR = stepwise nucleophilic reactant, SNTS1 = stepwise nucleophilic transition state 1, SNINT = stepwise nucleophilic intermediate, SNTS2 = stepwise nucleophilic transition state 2, SNP = stepwise nucleophilic product, CNR = concerted nucleophilic reactant, CNTS = concerted nucleophilic transition state, CNP = concerted nucleophilic product. All distances are in Å.

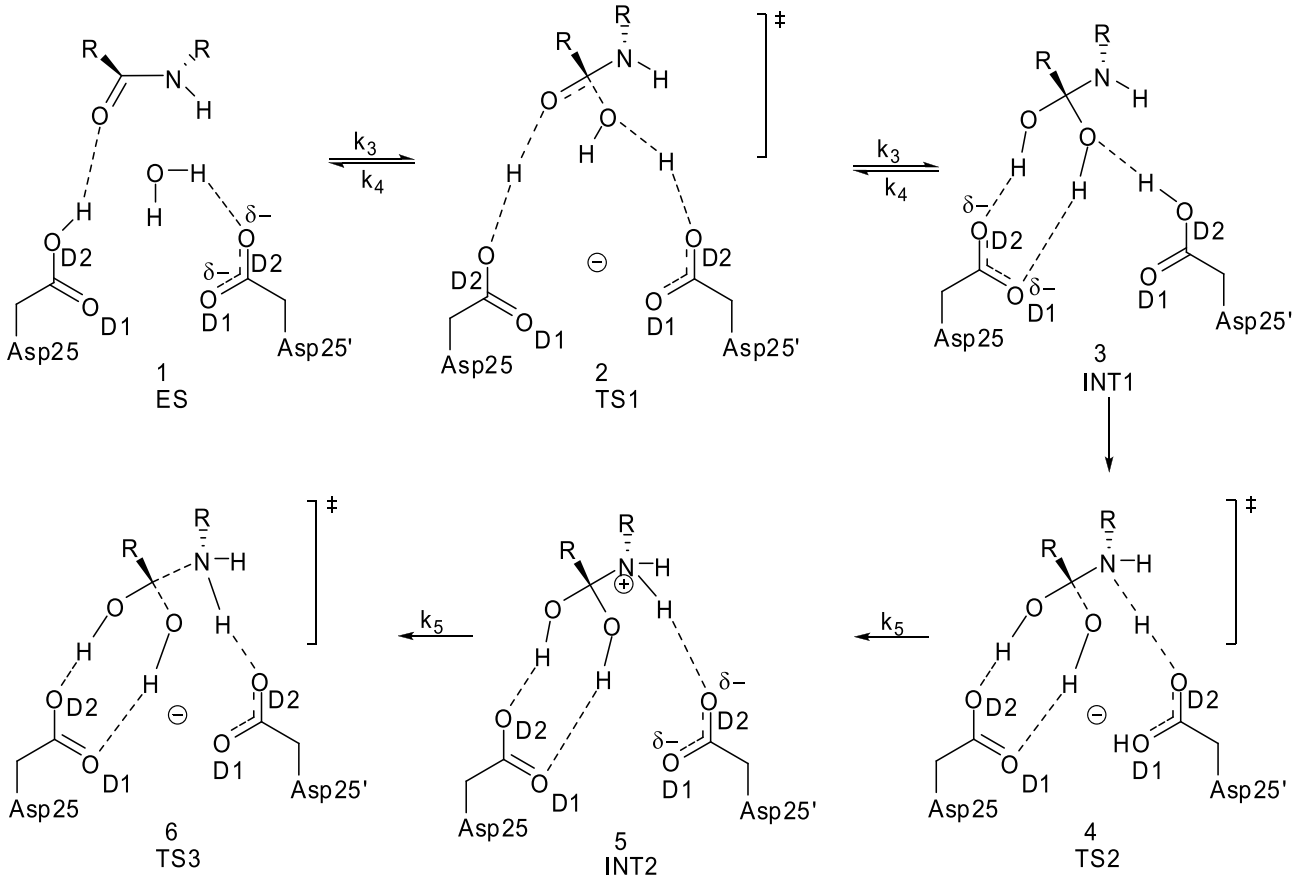
state (TS) leading to product (concerted nucleophilic reaction).[6,26]

In the general acid–base process, a catalytic water molecule acts as the nucleophile, as proposed by Fruton[30] and Suguna et al.[36] The mechanism is initiated with a nucleophilic attack on the carbonyl atom of the scissile unit by the water molecule at the active sites of the protease; this proceeds with the protonated Asp donating its hydrogen to the nitrogen of the scissile bond.[37] The unprotonated Asp ionizes the catalytic water by accepting its proton, and this process can either be a stepwise (Scheme 2) or a concerted (Scheme 3) reaction in which the former has been considered the most likely from the literature.[6,12,17,38] HIV-1 PR as well as their active mutants recognize and cleave the same Gag and Gag-Pol polyprotein precursors; hence, there remains a mutual structure–function relationship among them that should be exploited in existing and future drug inhibitors.[39]

Earlier theoretical studies that have focused on the catalytic mechanism of the HIV-1 PR utilized molecular mechanics (MM) and molecular dynamics (MD) modeling to explore the enzyme–substrate and enzyme–intermediate complexes.[26,42−44] One of these studies used a model system that focused mainly on the active site and the substrates,[26] while others only examined the active site with its closest surrounding groups.[43,45] The first notable limitation while exploring the mechanistic pathway required for substrate

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

Scheme 2. Proposed[a] Concerted General Acid-Base HIV-1 PR Mechanism via Asp 25 Protonation



[a]Structures along the pathway are indicated as follows: (1) enzyme−substrate complex; (2) water attack TS; (3) tetrahedral gem-diol intermediate; (4) scissile N-protonation TS; (5) protonated amide intermediate; and (6) cleavage of scissile C−N bond TS leading to a separated product complex. Redrawn from the literature.[39−41]

recognition and subsequent cleavage was the use of modified substrate sequences similar to the natural substrate. In some cases, six-alanine residues were used for the substrate sequence,[37] and in other investigations, one or more of the amino acids were substituted.[12,37,39] Another impediment can be associated with the conformational difference in the active site. Ribeiro et al.[46] reported that the energy barriers are dependent on the distances of the reactant to the transition state conformation. Also, the electronic charge is influenced by the residues at the active site and this contributes to the associated energy barriers.[46]

Over the years, improvements have been made in the use of theoretical methods to explore reaction processes. Density functional theory (DFT), which is a typical quantum mechanics (QM) method, has been considered suitable to elucidate mechanistic pathways and obtain results close to experimentally derived values.[47−50] The reactivity of the HIV-1 PR model was examined by Garrec et al. in 2011 to specifically test the accuracy of the latest DFT methods.[47] They suggested the combination of B3LYP[51,52]/6-31++G(d,p)[53,54] or B3LYP/6-311++G(d,p)[53,54] for geometry optimization, and the former theoretical level was adopted herein.

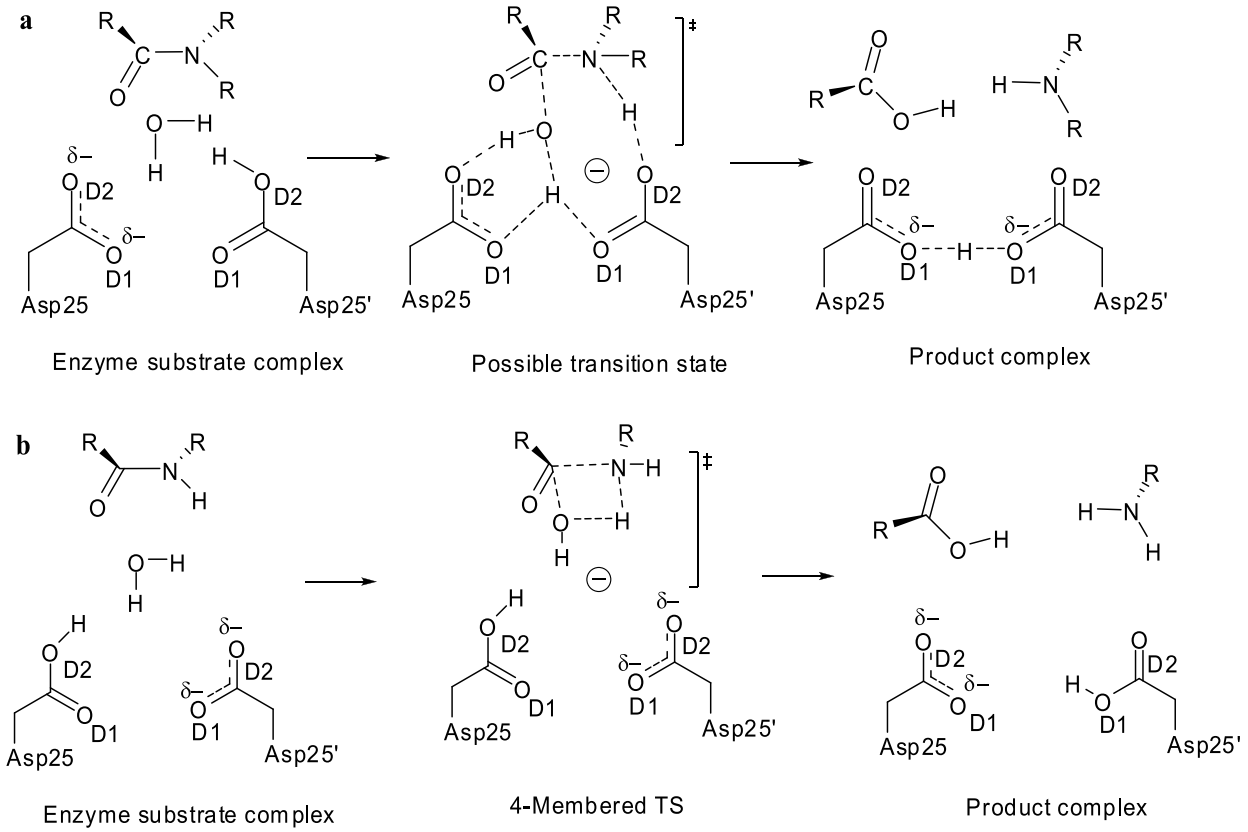
It is worth noting that two theoretical approaches[6,37] with regard to the one-step concerted mechanism for HIV PR cleavage of natural substrates are found in the literature (Scheme 3). In these studies, cyclic TS structures were employed, yielding activation free energies ($\Delta G^{\ddagger}$) of 30.0[6] and 43.5[37] kcal mol^{-1}. Most recently, the feasibility of the concerted general acid−base model involving two 6-membered TS, with either one or two water molecules, plus an acyclic TS (involving one water molecule) was investigated[55] (Scheme 4). The study[55] utilizes MD and the hybrid QM/MM approach in which experimentally comparable results ($\Delta G^{\ddagger}$ = 15.2−26.4 kcal mol^{-1}) were obtained. From experimental data, $\Delta G^{\ddagger}$ between 14.86 and 21.03 kcal mol^{-1} was estimated depending on the peptide sequence.[20,25,56−61]

Having observed a lower energy barrier in the one-step concerted acyclic TS reaction mechanism of the HIV-1 PR−substrate involving one water from our previous work (Scheme 4a),[55] we aimed to explore the preferential recognition and cleavage mechanism through TS modeling. Theoretical and experimental studies suggested that substrate modulation, substrate groove, conserved substrate shape, interdependence conformational adaptability of both enzyme and target, as well as conformational specificity are crucial for recognition of natural substrate by its HIV-1 PR.[17] For theoretical studies, the calculation of interaction and binding free energy parameters is required to investigate the mechanism of action between the substrate and the PR.[23,62−64]

To the best of our knowledge, on the basis of the latest literature review,[17] examining preferential cleavage of the natural substrates of HIV-1 PR via a concerted TS modeling using both MD and QM/MM methods has not been reported elsewhere. The one-step concerted acyclic general acid−base

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

Scheme 3. Proposed Concerted General Acid−Base HIV-1 PR Mechanism via (a) Asp25′ OD2 Protonation Based on the Jaskólski et al.[28] Experimental Hypothesis and (b) the 4-Membered Ring Model Redrawn from the Literature.[37]



mechanism requires monoprotonation of one aspartate in the active site of the enzyme (Scheme 4a).[55] The water molecule and both catalytic Asp (25/25′) are involved in the reaction mechanism, where the acidic Asp (protonated) loses its proton and becomes a base and the unprotonated Asp (base) becomes acidic by gaining a proton. The acyclic TS involved in this model for the elucidation of the concerted general acid−base pathway[55] proved to correlate well with experimental data.[17] This mechanism was therefore applied to obtain a better understanding of substrate recognition by calculating the activation free energies at different peptide bond regions (in addition to the scissile region) within the natural substrate sequence.

In order to ensure the calculation is not an artifact of a single enzyme−substrate system, two PRs, HIV-1 subtype B[65] and a popular mutant in South Africa (subtype C-SA)[66] complexed with MA-CA (Gag) and RH-IN (Gag-Pol) natural substrates were used. Likewise, the Gag and Gag-Pol polyprotein sites that are not normally cleaved (nonscissile domains) were also considered. Presented in Table S1 (from the Supporting Information) are the recognition sequences cleaved by HIV-1 PR. It is worth mentioning that the C-SA HIV-1 PR (Figure 1) differs structurally due to the eight-point amino acid residue mutations (R41K, L19I, T12S, H69K, I93L, I15V, L89M, and M36I) occurring from the subtype B PR.[67]

By means of concerted TS modeling, the specificity of HIV-1 PR for its natural targets was explored using constrained MD simulations and the hybrid QM/MM (Our Own N-layered Integrated molecular Orbital and molecular Mechanics, ONIOM)[68−71] algorithm. The modeled amino acid residues were varied from 6 to 10 units for detailed analysis and optimum recognition.

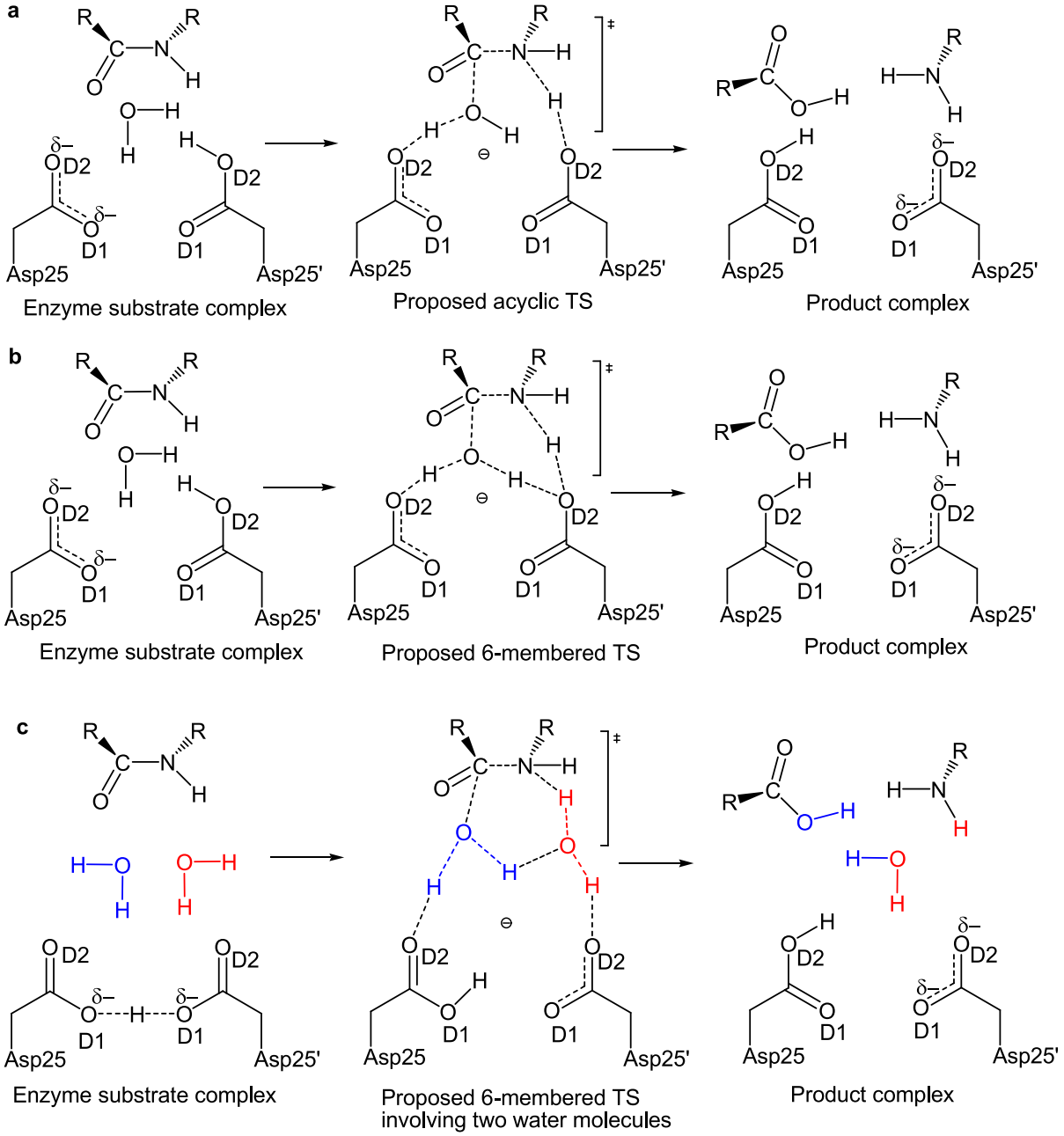
2. METHOD

2.1. The System Setup. Initial structure coordinates of the subtype B and C-SA HIV-1 PR were taken at 2.1 and 2.7 Å resolution from the Brookhaven Protein Data Bank[73] (PDB codes: 2P3B[65] and 3U71[66]) respectively. P3−P3′ amino acid sequence coordinates for substrates (MA-CA and RH-IN) were drawn from PDB structures 1KJ4 and 1KJH, respectively.[74] The MA-CA crystal complex was obtained at a 2.9 Å resolution,[74] and the remaining crystal complexes used (RH-IN and catalytic water) had a 2.0 Å resolution.[74,75]

For MA-CA, three cleavage domains were chosen from its substrate sequence: Val-Ser*Gln-Asn-**Tyr*Pro***Ile-Val-Gln-Asn (-**Tyr*Pro**- is the scissile bond, while -Ser*Gln- and -Pro*Ile- are the studied nonscissile domains of the same natural substrate). From the literature, a major discrepancy was discovered at the P2 position for RH-IN; in some cases, valine was utilized,[76,77] while other authors[23,78−80] adopted Ile. In this study, Ile was used because this is the most common amino acid sequence established[23,78−80] and the 3D structure[74] is available in the PDB Bank. Hence, for RH-IN, three cleavage domains were also chosen from this substrate: Ile-Arg*Lys-Ile-**Leu*Phe***Leu-Asp-Gly-Ile (-**Leu*Phe**- is the scissile bond, and -Phe*Leu- and -Arg*Lys- are the considered nonscissile regions within the same natural substrate). The activation energies for each of these cases were calculated for both subtype B and C-SA HIV-1 PR. First, a substrate model with six amino acid residues (P3−P3′) was used, and we later

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

Scheme 4. Proposed Concerted General Acid−Base Reaction Schemes for the Cleavage of Natural Substrate by HIV-1 PR: (a) an Acyclic Model; (b) a 6-Membered Cyclic Model; (c) a Two-Water Model Having a 6-Membered TS[55]



proceed to 10 peptide sequences (P5−P5′) to improve the (P3−P3′) recognition results for the RH-IN cleavage.

As water is an important factor for the HIV-1 PR proteolysis, a catalytic water molecule closest to the active site was taken from the PDB code 1LV1[75] and superimposed on both enzyme−substrate complexes considered; this is represented in Figure 2. This is necessary to ensure that the same pose is maintained in the binding sites of the structure complexes and serve as a starting structure for all of the investigated complexes. The position of the substrate with the catalytic water in the active site binds similarly to the reported HIV PR−inhibitor complexes,[81,82] and this was superimposed using PyMOL.[83] Thus, the crystallographic water molecules and ions present within the structures of the entire enzyme−substrate complexes were removed, leaving only the catalytic water molecule at the active site.

The pK_a values of the titratable amino acids were calculated using the empirical PROPKA,[84] and monoprotonation of Asp25′ OD2 (Scheme 4a) at physiological pH 7 was induced in the binding site, as supported from the literature[17,28] and our recent study[55] using GaussView.[85] Note that the standard pK_a values of ionizable groups can be shifted by protein environments.[86] The enzyme−substrate complexes with the Asp proton (H^+) were therefore positioned midway to the nitrogen of the scissile bond (Scheme 4a). This position has been shown to be stable energetically via MD simulations.[45]

The first results (Table 1) for the P3−P3′ RH-IN substrate revealed lower activation free energies for some of the nonscissile bonds. This prompted us to consider a longer

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

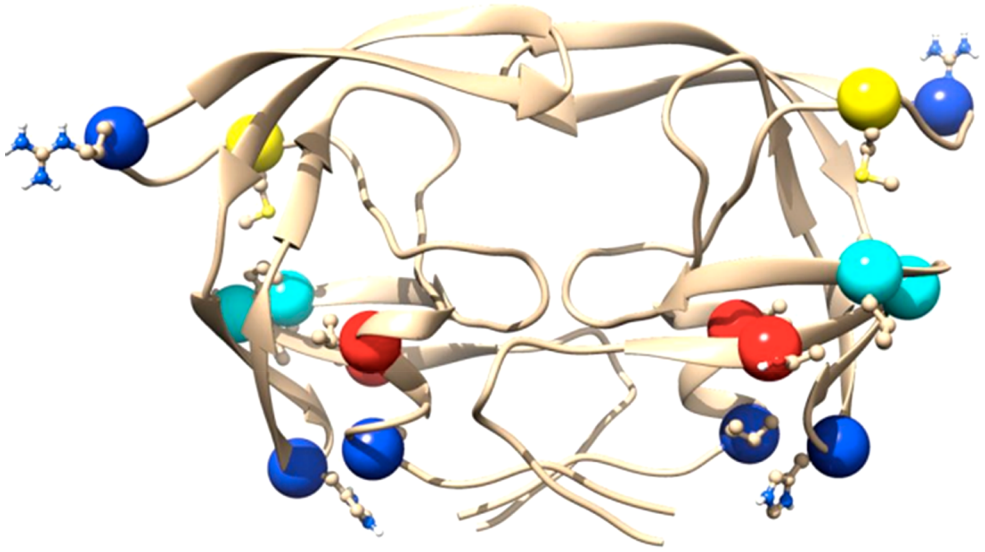


Figure 1. Homodimeric X-ray structure of subtype C-SA protease (PDB code: 3U71[66]) showing the positions occupied by the eight amino acids R41K, L19I, T12S, H69K, I93L, I15V, L89M, and M36I. This structure was created using UCSF chimera.[72]

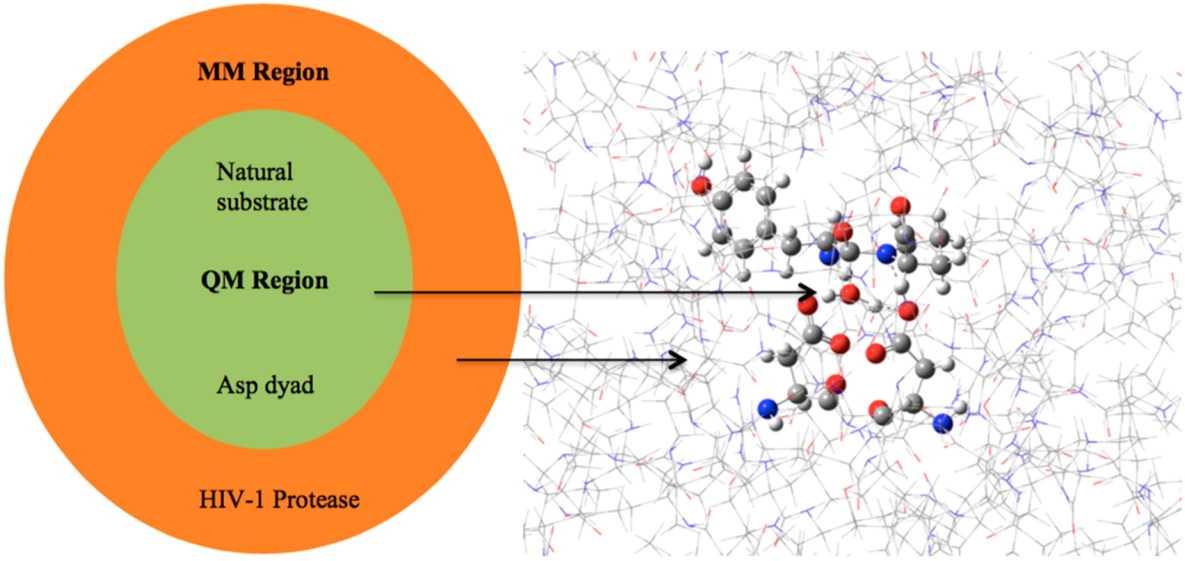


Figure 2. Schematic representation of the applied two-layered ONIOM model (B3LYP/6-31++G(d,p):AMBER) for the HIV-1 PR–MA-CA (P3–P3′) complex (acyclic TS with one water molecule).

substrate (P5–P5′) as well as perform a constrained MD simulation on the substrate to find the lowest energy enzyme–substrate complex as a starting point for the subsequent ONIOM transition state calculation. We postulated this approach may potentially provide improved recognition between the substrates and the enzyme prior to the TS calculation.

The MD simulation protocols for P5–P5′ HIV PR complexes were employed as described by Ribeiro et al.[46] This involves minimizations at 2500 cycles with the heating stage constraint to counteract the SHAKE failure that might result from an increase in temperature from 0 to 600 K. The constraint allowed the enzyme to be frozen, while the substrate was subjected to heating (MD script provided in the Supporting Information) up to 600 K. It was demonstrated before[87] that higher temperatures are required to overcome energy barriers of constrained molecules. Visual inspection of the movement of the substrate during the MD simulation revealed that extensive enzyme–substrate interactions were achieved. A stepwise equilibration was done in three stages; 10, 5, and 2, and the constraint force was reduced at each step. In the production stage, a 3–4 Å constraint distance between the OH of the water (Owat) with the carbonyl of the scissile bond (Csub) and the OH of the protonated Asp (OD2 Asp 25′) with the nitrogen of the scissile bond (Nsub) was fixed and run for 10 ns. Thereafter, a representative snapshot of the lowest energy conformation was obtained and an unconstrained MD calculation for 2 ns was performed at 300 K.

Results from the unconstrained MD for all complexes were superimposed with the enzymes' (B and C-SA) PDB structures, and these complexes were taken for subsequent ONIOM calculations. The acyclic concerted TS starting structures[55] were constructed using modred; interatomic distances were constrained using the respective geometries (Owat···Csub, Nsub···OD2 Asp25′, and Owat···OD2 Asp25′) and optimized to construct a suitable starting structure for an

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389–6400

Table 1. Relative Thermodynamic and Kinetic Parameters for the One-Step Concerted Acyclic TSs Involved in the Cleavage Mechanism of the P3−P3′ Natural Substrates for Scissile and Nonscissile Bonds by HIV-1 PRs Using ONIOM [B3LYP/6-31++G(d,p):AMBER][a]

MA-CA	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta\Delta G^{\ddagger}$
HIV-1 PR B					
Gln-Asn-Tyr*Pro-Ile-Val	−28.60	−23.49	−38.71	15.22	
Asn-Tyr-Pro*Ile-Val-Gln	−9.91	−10.49	−38.28	27.79	12.57
Gln-Val-Ser*Gln-Asn-Tyr	−14.43	−18.65	−41.25	22.60	7.38
HIV-1 PR C-SA					
Gln-Asn-Tyr*Pro-Ile-Val	−40.19	−39.32	−52.81	13.49	
Asn-Tyr-Pro*Ile-Val-Gln	−17.74	−21.98	−44.69	22.70	9.21
Gln-Val-Ser*Gln-Asn-Tyr	−32.74	−37.52	−55.12	17.60	4.11
RH-IN					
HIV-1 PR B					
Lys-Ile-Leu*Phe-Leu-Asp	−9.85	−10.60	−30.00	19.40	
Ile-Leu-Phe*Leu-Asp-Gly	−8.96	−9.70	−29.03	19.33	−0.07
Gly-Ile-Arg*Lys-Ile-Leu	−11.89	−12.92	−34.92	22.00	4.60
HIV-1 PR C-SA					
Lys-Ile-Leu*Phe-Leu-Asp	−25.72	−29.25	−47.28	18.03	
Ile-Leu-Phe*Leu-Asp-Gly	−27.71	−31.57	−49.54	17.97	−0.06
Gly-Ile-Arg*Lys-Ile-Leu	−28.47	−29.05	−49.79	20.74	2.71

[a]Values are reported in kcal mol^{-1} relative to the sum of separated reactants. The peptide sequences in bold are the natural substrate scissile bonds. $\Delta E^{\ddagger}$ = total energy, $\Delta H^{\ddagger}$ = enthalpy, $T\Delta S^{\ddagger}$ = entropy change in temperature, $\Delta G^{\ddagger}$ = activation free energy, $\Delta\Delta G^{\ddagger}$ = activation free energy difference.

unconstrained TS optimization. The active Asp25/25′ residues, catalytic water, as well as the substrate were partitioned to a high layer (QM/B3LYP[51,52]/6-31++G(d,p)[53,54]), while the remaining system was put to a lower layer (MM/AMBER[88,89]) for the ONIOM calculations. The electronic embedding from the ONIOM algorithm[90] was implemented to consider the QM/MM electric field, since the reaction center involves the transfer of protons during TS modeling.

2.2. Free Energy Calculations. To model the cleavage and formation of chemical bonds, electrons are included explicitly in the energy calculations,[76] and to achieve this in an entire enzyme, a hybrid method should be employed.[9] Hence, the enzyme was divided into two regions involving one classical model and a quantum mechanical (Figure 2) model. It is important for the latter to describe the chemistry of the reaction process correctly.[22] A two-layer ONIOM approach was thus applied to calculate the activation free energy of the catalyzed reaction. Preceding studies showed hybrid QM/MM methods are adequate for the HIV-1 PR catalytic mechanism[9,40,55,91] and the accuracy of the DFT/B3LYP method has been tested for HIV-1 PR reactivity.[47] All of the complex structures for obtaining starting optimized reactant states are prepared from energy minimization calculations using the Amber force field.[92] This allows the substrate and catalytic water to adjust to the binding site. This was followed by a full geometry optimization of the input structures in the gas phase at the B3LYP/6-31++G(d,p):AMBER level of theory. Vibrational frequencies[93] for the different unconstrained complexes were computed to characterize them as TS structures with one negative eigenvalue. In this study, GaussView 5.0.8[85] was used for both pre- and postprocessor graphic interface analyses with all calculations performed within the Gaussian 09[94] package.

Estimating the chemical quantities such as thermodynamics that arises from energy, entropy contribution, and enthalpy could be achieved computationally.[93] In the context of this study, the activation free energy, enthalpy, and temperature-dependent (at T = 298.15 K) change in entropy are denoted as $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$, and $T\Delta S^{\ddagger}$, respectively. Provided in eqs 1−3 are the activation free energy attained from frequency calculations for the studied polypeptide sequences involving the scissile bond and nonscissile bonds, as well as their difference within the enzyme−substrate system of both subtypes B and C-SA HIV-1 PR.

$$\Delta G^{\ddagger}_{\text{scissile}} = \Delta G_{\text{complex}} - \Delta G_{\text{protease}} - \Delta G_{\text{scissile}} \quad (1)$$

$$\Delta G^{\ddagger}_{\text{nonscissile}} = \Delta G_{\text{complex}} - \Delta G_{\text{protease}} - \Delta G_{\text{nonscissile}} \quad (2)$$

$$\Delta\Delta G^{\ddagger}_{\text{nonscissile}\rightarrow\text{scissile}} = \Delta G^{\ddagger}_{\text{nonscissile}} - \Delta G^{\ddagger}_{\text{scissile}} \quad (3)$$

3. RESULTS AND DISCUSSION

3.1. Recognition and Catalysis of Substrates with P3−P3′ Sequence by HIV-1 PRs. By means of modeling, the TS energy and recognition pattern of substrates by HIV-1 subtypes B and C-SA PRs was investigated. The reaction mechanism utilized here follows a concerted general acid−base mechanism where the unprotonated Asp25 acts as the base and the protonated Asp25′ acts as the acid in a one-step process (Scheme 4a). We have previously reported the acyclic concerted theoretical mechanism,[55] which corresponds well with reported experimental results. Recall that experimental $\Delta G^{\ddagger}$ values between 14.86 and 21.03 kcal mol^{-1} for natural substrates (and peptides) were deduced from the literature.[20,25,56−61]

In general, the results for the polypeptide sequences reveal that subtype C-SA has an increased rate of hydrolysis compared to subtype B for the hydrolysis of peptides (including natural substrates) by HIV-1 PR. All of the calculated activation free energies reported herein (Table 1) are in reasonable agreement with the experimental values. Specifically, the calculated value (15.22 kcal mol^{-1}) for the subtype B HIV-1 PR−MA-CA natural substrate scissile bond is in very good agreement with a reported experimental value of

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

Table 2. Relative Thermodynamic and Kinetic Parameters for the One-Step Concerted Acyclic TSs Involved in the Cleavage Mechanism of the P5−P5′ Natural Substrates for Scissile and Nonscissile Bonds by HIV-1 PRs Using ONIOM [B3LYP/6-31++G(d,p):AMBER][a]

RH-IN	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta\Delta G^{\ddagger}$
HIV-1 PR B					
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	−9.35	−10.62	−30.31	19.69	
Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp	−9.96	−10.80	−30.60	19.80	0.11
Ser-Ala-Gly-Ile-Arg*Lys-Ile-Leu-Phe-Leu	−11.99	−12.96	−37.25	24.29	4.60
HIV-1 PR C-SA					
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	−25.64	−29.34	−47.68	18.34	
Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp	−27.56	−30.59	−49.03	18.44	0.10
Ser-Ala-Gly-Ile-Arg*Lys-Ile-Leu-Phe-Leu	−29.17	−31.22	−52.26	21.04	2.70

[a]Values are reported in kcal mol^{-1} relative to the sum of separated reactants. The peptide sequences in bold are the natural substrate scissile bonds. $\Delta E^{\ddagger}$ = total energy, $\Delta H^{\ddagger}$ = enthalpy, $T\Delta S^{\ddagger}$ = entropy change in temperature, $\Delta G^{\ddagger}$ = activation free energy, $\Delta\Delta G^{\ddagger}$ = activation free energy difference.

15.69 kcal mol^{-1} [58] for MA-CA (Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln) hydrolysis by HIV-1 PR. Also, the calculated $\Delta G^{\ddagger}$ values of 18.03 and 17.97 kcal mol^{-1} for the cleavage of the HIV-1 PR subtype C-SA−RH-IN scissile domain and the Ile-Leu-Phe*Leu-Asp-Gly nonscissile region respectively agree with the observed experimental $\Delta G^{\ddagger}$ values of 18.24[20] and 17.60[57] kcal mol^{-1} for RH-IN hydrolysis by HIV-1 PR.

It is interesting to note that the scissile bond for MA-CA systems gave the lowest activation energies (compared to the nonscissile bonds) as expected. This is not the case for the RH-IN systems, suggesting the model substrates (P3−P3′) are likely too short for proper recognition. This observation could be akin to a recent (experimental and computational) study[64] in which it was suggested that an increase in the peptide sequence has a substantial effect on the protease selectivity of cleavage sites. In an earlier theoretical investigation by Perez et al.,[62] it was concluded that many nonscissile peptide sequences, despite their high affinity, do not cleave because they are suppressed inside the polyproteins.[62] Also, the recognition of substrate by HIV-1 PR is based on the geometric specificity of the exposed Gag and Gag-Pol polyprotein precursors.[25,62] However, in this present study whereby the nonscissile domains were modeled in proximity with the PRs, they did not necessarily yield lower activation free energies than the default scissile bonds (Table 1) despite being exposed to the enzymes' active region.

The overall mechanistic pathway is a synchronous process[55] with negative values for $\Delta E^{\ddagger}$, $\Delta H^{\ddagger}$, and $T\Delta S^{\ddagger}$ for the natural substrate and nonsubstrate transition states. The large negative values of the $T\Delta S^{\ddagger}$ indicate that the mechanistic process is largely driven by entropy contributions due to the concerted reaction process (Scheme 4a) for both HIV-1 PR B and C-SA−peptide complexes. On the basis of the accepted notion that a more negative entropy indicates greater restrictions of the inhibitor in the active site,[82,95] it is therefore likely that the calculated $T\Delta S^{\ddagger}$ values seem to have arisen from restriction of substrate movement in the active site.

The P3−P3′ calculated energetics also revealed that subtype C-SA has more negative $\Delta E^{\ddagger}$ and $\Delta H^{\ddagger}$ values than subtype B. To further understand the specificity of the PRs for the peptide sequences, activation free energy profiles were computed for both scissile and nonscissile regions of the Gag and Gag-Pol polyprotein segments (Table S1).

3.2. MA-CA Substrate Specificity and Analysis of the Complexes Involving the P3−P3′ Unit. The subtype B−MA-CA natural substrate complexes have activation free energy differences ($\Delta\Delta G^{\ddagger}$) of approximately 12.57 kcal mol^{-1} (Asn-Tyr-Pro*Ile-Val-Gln) and 7.38 kcal mol^{-1} (Gln-Val-Ser*Gln-Asn-Tyr) compared to the scissile domain (Gln-Asn-**Tyr*Pro**-Ile-Val). This outcome is feasible, as observed by Boross et al.[25] where HIV-1 substrates show preference for small residues such as Asn or Cys at the P2 position and Ile or Val at the P2′ position for HIV-1 PR. For the Tyr-Pro*Ile-Val nonscissile sequence with Val at the P2′ position, a higher activation free energy of 27.79 kcal mol^{-1} was computed. This can be attributed to the presence of the large hydrophobic chains of proline at the P1 position, while the nonscissile bond with small residues (Ser*Gln) has a better affinity. Although subtype B PR correctly recognizes and cleaves the nonscissile peptide bonds, the effect of the Pro side chain in the Asn-Tyr-Pro*Ile-Val-Gln sequence gave a $\Delta\Delta G^{\ddagger}$ value of 5.19 kcal mol^{-1} compared to the scissile bond domain (Gln-Val-Ser*Gln-Asn-Tyr). It is quite fascinating to note that the calculated $\Delta G^{\ddagger}$ value (15.22 kcal mol^{-1}) for the subtype B HIV-1 PR−MA-CA natural substrate scissile bond is in very good agreement with a reported experimental value of 15.69 kcal mol^{-1} for MA-CA (Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln) hydrolysis by HIV-1 PR.[58]

For the subtype C-SA−MA-CA complex, a $\Delta G^{\ddagger}$ value of 13.49 kcal mol^{-1} was obtained for the cleavage of the scissile bond (Gln-Asn-**Tyr*Pro**-Ile-Val), with a decreased $\Delta\Delta G^{\ddagger}$ value of −1.73 kcal mol^{-1} in comparison with the subtype B−MA-CA complex. The rate of hydrolysis for subtype C-SA−MA-CA is faster, and this could be a result of the mutations that occur in the protease, which also contributes to the less effective inhibition of the FDA approved drugs.[81,96] This lower $\Delta G^{\ddagger}$ value could also be related to the HIV-1 PR preference for hydrophobic amino acid side chains (Tyr*Pro) at the P1 and P1′ positions of the natural target. Furthermore, the nonscissile bond domain that contains Pro*Ile has a $\Delta G^{\ddagger}$ value of 22.70 kcal mol^{-1}, that is 9.21 kcal mol^{-1} higher than its natural substrate and 5.09 kcal mol^{-1} lower than the subtype B complex for the same cleavage site (Asn-Tyr-Pro*Ile-Val-Gln).

These results suggested that the proposed concerted general acid−base acyclic cleavage mechanism (Scheme 4a) involving HIV-1 PR and its substrates (plus nonsubstrates) accurately predicts the recognition and specificity phenomena of the HIV-1 PR enzyme for its MA-CA substrates. Also, the recognition of the studied peptides is dependent on their geometry orientation, which was proposed earlier[78] plus their amino acid sequence length as recently reported.[64] From our observation, the model RH-IN (P3−P3′) system seems too

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

short for effective recognition; hence, we decided to study a P5−P5′ model for this substrate.

3.3. RH-IN Substrate Specificity and Analysis of the Complexes Involving the P5−P5′ Unit. The relative thermodynamic parameters for the HIV-1 PR−RH-IN catalysis of the P5−P5′ substrates are presented in Table 2. In all cases, the scissile bonds exhibited lower activation energies (albeit marginal) than the nonscissile bonds. An activation free energy of 19.69 kcal mol^{-1} was observed for the scissile bond cleavage of the subtype B−RH-IN natural substrate complex, which is 0.11 kcal mol^{-1} lower than the considered nonscissile domain (Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp). Likewise, the subtype C-SA−RH-IN natural substrate system was observed to follow this trend with a $\Delta\Delta G^{\ddagger}$ difference of 0.10 and 2.70 kcal mol^{-1} compared to the nonscissile peptide sequences Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp and Ser-Ala-Gly-Ile-Arg*Lys-Ile-Leu-Phe-Leu, respectively (Table 2).

For both proteases, HIV-1 PR complexed with the nonscissile domain Ser-Ala-Gly-Ile-Arg*Lys-Ile-Leu-Phe-Leu, activation free energies of 24.29 and 21.04 kcal mol^{-1} were obtained, respectively. Although a slight increase for the activation energy of the scissile bonds was observed in the P5−P5′ calculation, the longer peptide sequences provide better recognition.

4. CONCLUSION

In this study, we have investigated the mechanistic route of substrate recognition between subtype B and C-SA HIV-1 PR, using constraint MD simulations and a two-layered ONIOM B3LYP/6-31++G(d,p) method, and found that the substrate pattern recognition is similar for both wild type and mutated enzymes. The cleavage mechanism follows a concerted acyclic TS model leading to product and activation free energies that are in reasonable agreement with experimental values for peptide hydrolysis by HIV-1 PR. It needs to be emphasized that the recognition specificity of the enzyme depends on the Gag and Gag-Pol sequence as well as the number of amino acid residues in the substrate polypeptide sequence. The results obtained herein demonstrate that optimum recognition depends on both the length of the amino acid sequence and the structural details of the substrates. For example, the P3−P3′ MA-CA substrate provides suitable recognition, while a P5−P5′ sequence is required for effective recognition of the RH-IN substrates.

It is evident from the present results that the rate of hydrolysis for HIV-1 subtype C-SA PR is faster for both MA-CA and RH-IN (scissile-bond) complexes compared to the subtype B HIV-1 PR−substrate (scissile bond) complexes. This phenomenon could justify the inhibition of the enzyme in which the approved FDA HIV-1 PR inhibitors relatively inhibit the subtype B better than the subtype C-SA PR as reported in our previous work.[81] On the basis of accepted theory that the HIV-1 PR inhibitors mimic the transition state analogue of the enzyme-catalyzed reaction, we therefore propose that future potent inhibitors could be developed to mimic the RH-IN substrate for both subtype B and C-SA HIV-1 PR on the basis of its mechanism of reaction. Note that a higher activation free energy has been associated with a diminished hydrolytic ability.[49,50] Future research would be directed toward increasing the activation energy of these natural substrates through electronic effects, resulting in higher activation free energy. If that can be achieved, new natural substrate based inhibitors may be unveiled.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.9b02207.

Table S1 showing the nine recognition sequences cleaved by HIV-1 PR and the distance restraint/constraint molecular dynamics script (PDF)

ONIOM output files of the TS structures are provided in PDB format (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kruger@ukzn.ac.za. Fax: +27312601845.

ORCID

Bahareh Honarparvar: 0000-0001-9005-2282

Hendrik G. Kruger: 0000-0003-0606-2053

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors appreciate College of Health Sciences, University of KwaZulu-Natal, Asphen Pharmacare, Medical Research Council, and the National Research Foundation (all in South Africa) for financial support. We are also grateful to the Center for High Performance Computing (www.chpc.ac.za) for computational resources.

■ REFERENCES

(1) Barkan, D. T.; Hostetter, D. R.; Mahrus, S.; Pieper, U.; Wells, J. A.; Craik, C. S.; Sali, A. Prediction of Protease Substrates Using Sequence and Structure Features. *Bioinformatics* **2010**, *26*, 1714−1722.

(2) Pethe, M. A.; Rubenstein, A. B.; Khare, S. D. Data-Driven Supervised Learning of a Viral Protease Specificity Landscape from Deep Sequencing and Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 168−176.

(3) Feinberg, M. B. Changing the Natural History of HIV Disease. *Lancet* **1996**, *348*, 239−246.

(4) Shehu-Xhilaga, M.; Oelrichs, R. Basic HIV Virology. *HIV Management in Australasia* **2009**, 9−18.

(5) Hsu, I.-N.; Delbaere, L. T.; James, M. N.; Hofmann, T. Penicillopepsin from Penicillium Janthinellum Crystal Structure at 2.8 Å and Sequence Homology with Porcine Pepsin. *Nature* **1977**, *266*, 140.

(6) Park, H.; Suh, J.; Lee, S. Ab Initio Studies on the Catalytic Mechanism of Aspartic Proteinases: Nucleophilic Versus General Acid/General Base Mechanism. *J. Am. Chem. Soc.* **2000**, *122*, 3901−3908.

(7) Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R.; Scolnick, E. M.; Sigal, I. S. Active Human Immunodeficiency Virus Protease is Required for Viral Infectivity. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4686−4690.

(8) Weber, I. T.; Cavanaugh, D. S.; Harrison, R. W. Models of HIV-1 Protease with Peptides Representing Its Natural Substrates. *J. Mol. Struct.: THEOCHEM* **1998**, *423*, 1−12.

(9) Chatfield, D. C.; Eurenius, K. P.; Brooks, B. R. HIV-1 Protease Cleavage Mechanism: A Theoretical Investigation Based on Classical MD Simulation and Reaction Path Calculations Using a Hybrid QM/MM Potential. *J. Mol. Struct.: THEOCHEM* **1998**, *423*, 79−92.

(10) Veronese, F. D.; DeVico, A. L.; Copeland, T. D.; Oroszlan, S.; Gallo, R. C.; Sarngadharan, M. Characterization of gp41 as the Transmembrane Protein Coded by the HTLV-III/LAV Envelope Gene. *Science* **1985**, *229*, 1402−1405.

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

(11) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr Calculation of Solvation Free Energies Using a Density Functional/Molecular Dynamics Coupled Potential. *J. Phys. Chem.* **1993**, *97*, 11868−11870.
(12) Trylska, J.; Grochowski, P.; McCammon, J. A. The Role of Hydrogen Bonding in the Enzymatic Reaction Catalyzed by HIV-1 Protease. *Protein Sci.* **2004**, *13*, 513−528.
(13) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. H. Ionization States of the Catalytic Residues in HIV-1 Protease. *Nat. Struct. Biol.* **1996**, *3*, 946−950.
(14) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab Initio Molecular Dynamics-Based Assignment of the Protonation State of Pepstatin A/HIV-1 Protease Cleavage Site. *J. Am. Chem. Soc.* **2001**, *123*, 8730−8737.
(15) Northrop, D. B. Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Acc. Chem. Res.* **2001**, *34*, 790−797.
(16) Mcgee, T. D.; Edwards, J.; Roitberg, A. E. pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State. *J. Phys. Chem. B* **2014**, *118*, 12577−12585.
(17) Lawal, M. M.; Sanusi, Z. K.; Govender, T.; Maguire, G. E. M.; Honarparvar, B.; Kruger, H. G. From Recognition to Reaction Mechanism: An Overview on the Interactions between HIV-1 Protease and Its Natural Targets. *Curr. Med. Chem.* **2018**, *26*, 1−34.
(18) Navia, M. A.; Fitzgerald, P. M.; McKeever, B. M.; Leu, C.-T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1. *Nature* **1989**, *337*, 615.
(19) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F. X-Ray Analysis of HIV-1 Proteinase at 2.7 Å Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342*, 299.
(20) Hyland, L. J.; Tomaszek, T. A., Jr; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454−8463.
(21) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543−585.
(22) Polgar, L.; Szeltner, Z.; Boros, I. Substrate-Dependent Mechanisms in the Catalysis of Human Immunodeficiency Virus Protease. *Biochemistry* **1994**, *33*, 9351−9357.
(23) Özen, A.; Haliloğlu, T.; Schiffer, C. A. Dynamics of Preferential Substrate Recognition in HIV-1 Protease: Redefining the Substrate Envelope. *J. Mol. Biol.* **2011**, *410*, 726−744.
(24) Tözsér, J.; Bláha, I.; Copeland, T. D.; Wondrak, E. M.; Oroszlan, S. Comparison of the HIV-1 and HIV-2 Proteinases Using Oligopeptide Substrates Representing Cleavage Sites in Gag and Gag-Pol Polyproteins. *FEBS Lett.* **1991**, *281*, 77−80.
(25) Boross, P.; Bagossi, P.; Copeland, T. D.; Oroszlan, S.; Louis, J. M.; Tozser, J. Effect of Substrate Residues on the P2' Preference of Retroviral Proteinases. *Eur. J. Biochem.* **1999**, *264*, 921−929.
(26) Liu, H. Y.; MullerPlathe, F.; vanGunsteren, W. F. A. Combined Quantum/Classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease. *J. Mol. Biol.* **1996**, *261*, 454−469.
(27) Antonov, V. K.; Ginodman, L. M.; Rumsh, L. D.; Kapitannikov, Y. V.; Barshevskaya, T. N.; Yavashev, L. P.; Gurova, A. G.; Volkova, L. I. Studies on the Mechanisms of Action of Proteolytic Enzymes Using Heavy Oxygen Exchange. *Eur. J. Biochem.* **1981**, *117*, 195−200.
(28) Jaskólski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B.; Wlodawer, A. Structure at 2.5-. Ang. Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-Based Inhibitor. *Biochemistry* **1991**, *30*, 1600−1609.
(29) Chatfield, D. C.; Brooks, B. R. HIV-1 Protease Cleavage Mechanism Elucidated with Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1995**, *117*, 5561−5572.
(30) Fruton, J. S. The Mechanism of the Catalytic Action of Pepsin and Related Acid Proteinases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **2006**, *44*, 1−36.
(31) Antonov, V.; Ginodman, L.; Kapitannikov, Y. V.; Barshevskaya, T.; Gurova, A.; Rumsh, L. Mechanism of Pepsin Catalysis: General Base Catalysis by the Active-Site Carboxylate Ion. *FEBS Lett.* **1978**, *88*, 87−90.
(32) Bott, R.; Subramanian, E.; Davies, D. Three-Dimensional Structure of the Complex of the Rhizopus Chinensis Carboxyl Proteinase and Pepstatin at 2.5. Ang. Resolution. *Biochemistry* **1982**, *21*, 6956−6962.
(33) Polgar, L. *Mechanisms of Protease Action*; CRC Press: 1989.
(34) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S.; Wlodawer, A. X-Ray Crystallographic Structure of a Complex between a Synthetic Protease of Human Immunodeficiency Virus 1 and a Substrate-Based Hydroxyethylamine Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8805−8809.
(35) Rodriguez, E. J.; Angeles, T. S.; Meek, T. D. Use of N-15 Kinetic Isotope Effects to Elucidate Details of the Chemical Mechanism of Human Immunodeficiency Virus-1 Protease. *Biochemistry* **1993**, *32*, 12380−12385.
(36) Suguna, K. A.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase from Rhizopus Chinensis: Implications for a Mechanism of Action. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7009−7013.
(37) Krzeminska, A.; Moliner, V.; Swiderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J. Am. Chem. Soc.* **2016**, *138*, 16283−16298.
(38) Trylska, J.; Bala, P.; Geller, M.; Grochowski, P. Molecular Dynamics Simulations of the First Steps of the Reaction Catalyzed by HIV-1 Protease. *Biophys. J.* **2002**, *83*, 794−807.
(39) Kipp, D. R.; Hirschi, J. S.; Wakata, A.; Goldstein, H.; Schramm, V. L. Transition States of Native and Drug-Resistant HIV-1 Protease Are the Same. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 6543−6548.
(40) Altoè, P.; Stenta, M.; Bottoni, A.; Garavelli, M. A Tunable QM/MM Approach to Chemical Reactivity, Structure and Physico-Chemical Properties Prediction. *Theor. Chem. Acc.* **2007**, *118*, 219−240.
(41) Trylska, J.; Grochowski, P.; Geller, M. Parameterization of the Approximate Valence Bond (AVB) Method to Describe Potential Energy Surface in the Reaction Catalyzed by HIV-1 Protease. *Int. J. Quantum Chem.* **2001**, *82*, 86−103.
(42) Tropsha, A.; Hermans, J. Application of Free Energy Simulations to the Binding of a Transition-State-Analogue Inhibitor to HTV Protease. *Protein Eng., Des. Sel.* **1992**, *5*, 29−33.
(43) Goldblum, A.; Glick, M.; Rayan, A. Determining Proton Positions in an Enzyme-Inhibitor Complex Is a First Step for Theoretical Mechanistic Studies of Aspartic Proteinases. *Theor. Chem. Acc.* **1993**, *85*, 231−247.
(44) Weber, I. T.; Harrison, R. W. Molecular Mechanics Calculations on HIV-1 Protease with Peptide Substrates Correlate with Experimental Data. *Protein Eng., Des. Sel.* **1996**, *9*, 679−690.
(45) Beveridge, A.; Heywood, G. A. Quantum Mechanical Study of the Active Site of Aspartic Proteinases. *Biochemistry* **1993**, *32*, 3325−3333.
(46) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Rarnos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catal.* **2015**, *5*, 5617−5626.
(47) Garrec, J.; Sautet, P.; Fleurat-Lessard, P. Understanding the HIV-1 Protease Reactivity with DFT: What Do We Gain from Recent Functionals? *J. Phys. Chem. B* **2011**, *115*, 8545−8558.
(48) Lawal, M. M.; Govender, T.; Maguire, G. E.; Honarparvar, B.; Kruger, H. G. Mechanistic Investigation of the Uncatalyzed Esterification Reaction of Acetic Acid and Acid Halides with Methanol: A DFT Study. *J. Mol. Model.* **2016**, *22*, 235−246.
(49) Fakhar, Z.; Govender, T.; Lamichhane, G.; Maguire, G. E.; Kruger, H. G.; Honarparvar, B. Computational Model for the Acylation Step of the B-Lactam Ring: Potential Application for L,

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

D-Transpeptidase 2 in *Mycobacterium Tuberculosis*. *J. Mol. Struct.* **2017**, *1128*, 94−102.
(50) Lawal, M. M.; Govender, T.; Maguire, G. E.; Kruger, H. G.; Honarparvar, B. DFT Study of the Acid-Catalyzed Esterification Reaction Mechanism of Methanol with Carboxylic Acid and Its Halide Derivatives. *Int. J. Quantum Chem.* **2018**, *118*, e25497−e25508.
(51) Lee, C.; Yang, W.; Parr, R. G. Development of Colle-Salvetti Correlation-Energy Formula into a Functional of Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(52) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(53) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* Basis Set for Atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223−1229.
(54) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-Row Atoms. *J. Comput. Chem.* **2001**, *22*, 976−984.
(55) Lawal, M. M.; Sanusi, Z. K.; Govender, T.; Tolufashe, G. F.; Maguire, G. E.; Honarparvar, B.; Kruger, H. G. Unraveling the Concerted Catalytic Mechanism of the Human Immunodeficiency Virus Type 1 (HIV-1) Protease: A Hybrid QM/MM Study. *Struct. Chem.* **2019**, *30*, 409−417.
(56) Hyland, L. J.; Tomaszek, T. A., Jr; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. Human Immunodeficiency Virus-1 Protease. 1. Initial Velocity Studies and Kinetic Characterization of Reaction Intermediates by ^{18}O Isotope Exchange. *Biochemistry* **1991**, *30*, 8441−8453.
(57) Rodriguez, E. J.; Debouck, C.; Deckman, I. C.; Abu-Soud, H.; Raushel, F. M.; Meek, T. D. Inhibitor Binding to the Phe53trp Mutant of HIV-1 Protease Promotes Conformational Changes Detectable by Spectrofluorometry. *Biochemistry* **1993**, *32*, 3557−3563.
(58) Maschera, B.; Darby, G.; Palu, G.; Wright, L. L.; Tisdale, M.; Myers, R.; Blair, E. D.; Furfine, E. S. Human Immunodeficiency Virus - Mutations in the Viral Protease That Confer Resistance to Saquinavir Increase the Dissociation Rate Constant of the Protease-Saquinavir Complex. *J. Biol. Chem.* **1996**, *271*, 33231−33235.
(59) Tozser, J.; Bagossi, P.; Weber, I. T.; Copeland, T. D.; Oroszlan, S. Comparative Studies on the Substrate Specificity of Avian Myeloblastosis Virus Proteinase and Lentiviral Proteinases. *J. Biol. Chem.* **1996**, *271*, 6781−6788.
(60) Feher, A.; Weber, I. T.; Bagossi, P.; Boross, P.; Mahalingam, B.; Louis, J. M.; Copeland, T. D.; Torshin, I. Y.; Harrison, R. W.; Tozser, J. Effect of Sequence Polymorphism and Drug Resistance on Two HIV-1 Gag Processing Sites. *Eur. J. Biochem.* **2002**, *269*, 4114−4120.
(61) Kipp, D. R.; Silva, R. G.; Schramm, V. L. Mass-Dependent Bond Vibrational Dynamics Influence Catalysis by HIV-1 Protease. *J. Am. Chem. Soc.* **2011**, *133*, 19358−19361.
(62) Perez, M.; Fernandes, P.; Ramos, M. Substrate Recognition in HIV-1 Protease: A Computational Study. *J. Phys. Chem. B* **2010**, *114*, 2525−2532.
(63) Laco, G. S. HIV-1 Protease Substrate-Groove: Role in Substrate Recognition and Inhibitor Resistance. *Biochimie* **2015**, *118*, 90−103.
(64) Laco, G. S. Retroviral Proteases: Correlating Substrate Recognition with Both Selected and Native Inhibitor Resistance. *JMolBiochem* **2017**, *6*, 45−63.
(65) Kempf, D. J.; Norbeck, D. W.; Codacovi, L.; Wang, X. C.; Kohlbrenner, W. E.; Wideburg, N. E.; Paul, D. A.; Knigge, M. F.; Vasavanonda, S. Structure-Based, C2 Symmetric Inhibitors of HIV Protease. *J. Med. Chem.* **1990**, *33*, 2687−2689.
(66) Naicker, P.; Achilonu, I.; Fanucchi, S.; Fernandes, M.; Ibrahim, M. A.; Dirr, H. W.; Soliman, M. E.; Sayed, Y. Structural Insights into the South African HIV-1 Subtype C Protease: Impact of Hinge Region Dynamics and Flap Flexibility in Drug Resistance. *J. Biomol. Struct. Dyn.* **2013**, *31*, 1370−1380.
(67) Mosebi, S.; Morris, L.; Dirr, H. W.; Sayed, Y. Active-Site Mutations in the South African Human Immunodeficiency Virus Type 1 Subtype C Protease Have a Significant Impact on Clinical Inhibitor Binding: Kinetic and Thermodynamic Study. *J. Virol.* **2008**, *82*, 11476−11479.
(68) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1−21.
(69) Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital+ Molecular Orbital) Method. *J. Comput. Chem.* **2000**, *21*, 1419−1432.
(70) Morokuma, K. New Challenges in Quantum Chemistry: Quests for Accurate Calculations for Large Molecular Systems. *Philos. Trans. Royal Soc. A* **2002**, *360*, 1149−1164.
(71) Chung, L. W.; Sameera, W.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.
(72) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(73) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(74) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. Substrate Shape Determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes. *Structure* **2002**, *10*, 369−381.
(75) Kumar, M.; Kannan, K.; Hosur, M.; Bhavesh, N. S.; Chatterjee, A.; Mittal, R.; Hosur, R. Effects of Remote Mutation on the Autolysis of HIV-1 PR: X-Ray and NMR Investigations. *Biochem. Biophys. Res. Commun.* **2002**, *294*, 395−401.
(76) Pettit, S. C.; Clemente, J. C.; Jeung, J. A.; Dunn, B. M.; Kaplan, A. H. Ordered Processing of the Human Immunodeficiency Virus Type 1 Gagpol Precursor Is Influenced by the Context of the Embedded Viral Protease. *J. Virol.* **2005**, *79*, 10601−10607.
(77) Liu, Z.; Wang, Y.; Brunzelle, J.; Kovari, I. A.; Kovari, L. C. Nine Crystal Structures Determine the Substrate Envelope of the MDR HIV-1 Protease. *Protein J.* **2011**, *30*, 173−183.
(78) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How Does a Symmetric Dimer Recognize an Asymmetric Substrate? A Substrate Complex of HIV-1 Protease. *J. Mol. Biol.* **2000**, *301*, 1207−1220.
(79) Sadiq, S. K.; Wright, D.; Watson, S. J.; Zasada, S. J.; Stoica, I.; Coveney, P. V. Automated Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases. *J. Chem. Inf. Model.* **2008**, *48*, 1909−1919.
(80) Costa, M. G. S.; Benetti-Barbosa, T. G.; Desdouits, N.; Blondel, A.; Bisch, P. M.; Pascutti, P. G.; Batista, P. R. Impact of M36I Polymorphism on the Interaction of HIV-1 Protease with Its Substrates: Insights from Molecular Dynamics. *BMC Genomics* **2014**, *15*, S5.
(81) Sanusi, Z. K.; Govender, T.; Maguire, G. E. M.; Maseko, S. B.; Lin, J.; Kruger, H. G.; Honarparvar, B. Investigation of the Binding Free Energies of FDA Approved Drugs against Subtype B and C-SA HIV PR: ONIOM Approach. *J. Mol. Graphics Modell.* **2017**, *76*, 77−85.
(82) Sanusi, Z. K.; Govender, T.; Maguire, G. E.; Maseko, S. B.; Lin, J.; Kruger, H. G.; Honarparvar, B. An Insight to the Molecular Interactions of the FDA Approved HIV PR Drugs against L38l↑ N↑ L PR Mutant. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 459−471.
(83) DeLano, W. L. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography* **2002**, *40*, 82−92.
(84) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Struct., Funct., Genet.* **2005**, *61*, 704−721.
(85) Dennington, R. D.; Keith, T. A.; Millam, J. M. *GaussView*, version 5.0.8; Gaussian, Inc.: Wallingford, CT, 2009.
(86) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-Dependent Properties of Proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

(87) Honarparvar, B.; Govender, T.; Maguire, G. E. M.; Soliman, M. E. S.; Kruger, H. G. Integrated Approach to Structure-Based Enzymatic Drug Design: Molecular Modeling, Spectroscopy, and Experimental Bioactivity. *Chem. Rev.* **2014**, *114*, 493−537.
(88) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(89) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wires Comput. Mol. Sci.* **2013**, *3*, 198−210.
(90) Vreven, T.; Morokuma, K. Hybrid Methods: ONIOM (QM:MM) and QM/MM. *Annu. Rep. Comput. Chem.* **2006**, *2*, 35−51.
(91) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(92) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(93) Ochterski, J. W. *Vibrational Analysis in Gaussian*; help@gaussian.com; 1999.
(94) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, rev D.01; Gaussian, Inc.: Wallingford, CT, 2009.
(95) Ryde, U. A. Fundamental View of Enthalpy-Entropy Compensation. *MedChemComm* **2014**, *5*, 1324−1336.
(96) Maseko, S. B.; Padayachee, E.; Govender, T.; Kruger, G.; Maguire, G. E. M.; Lin, J.; Sayed, Y. I36T↑T Mutation in South African Subtype C (C-SA) HIV-1 Protease Significantly Alters Protease-Drug Interactions. *Biol. Chem.* **2017**, *398*, 1109−1117.

DOI: 10.1021/acs.jpcb.9b02207
J. Phys. Chem. B 2019, 123, 6389−6400

SUPPORTING INFORMATION

Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates

Zainab K. Sanusi,[a] Monsurat M. Lawal,[a] Thavendran Govender,[a] Glenn E. M. Maguire,[a, b] Bahareh Honarparvar,[a] and Hendrik G. Kruger.[a,*]

[a]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

[b]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.

*Corresponding author: Prof. H. G. Kruger; kruger@ukzn.ac.za, Fax: +27312601845, Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

Section S1

Table S1: The nine recognition sequences cleaved by HIV-1 PR.[1-2]

Cleavage sites in Gag	Cleavage domain
Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-Asn	MA-CA
Lys-Ala-Arg-Val-Leu*Ala-Glu-Ala-Met-Ser	CA-p2
Pro-Ala-Thr-Ile-Met*Met-Gln-Arg-Gly-Asn	p2-NC
Glu-Arg-Gln-Ala-Asn*Phe-Leu-Gly-Lys-Ile	NC-p1
Arg-Pro-Gly-Asn-Phe*Leu-Gln-Ser-Arg-Pro	p1-p6
Cleavage sites in Gag-Pol	
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	RH-IN
Val-Ser-Phe-Asn-Phe*Pro-Gln-Ile-Thr-Leu	TF-PR
Cys-Thr-Leu-Asn-Phe*Pro-Ile-Ser-Pro-Ile	PR-RT
Gly-Ala-Glu-Thr-Phe*Tyr-Val-Asp-Gly-Ala	RT-RH

The asterisk (*) denotes the cleavage sites and are named after the proteins that are released. Matrix-capsid; MA-CA, capsid-p2; CA-p2, p2-nucleopsid; p2-NC, nucleopsid-p1; NC-p1, *trans* frame peptide-protease; TF-PR, protease-reverse transcriptase; PR-RT, reverse transcriptase-RNAseH; RT-RH, RNAseH-integrase; RH-IN.

Section S2

Distance Restraint/Constraint Molecular Dynamics[3]

1. Create a pdb file using the prmtop/inpcrd files of the simulation you are intended to do. This .pdb file will order the atoms according to prmtop file. This is a very crucial step!

ambpdb -aatm -p prmtop <inpcrd> structure.aatm.pdb

ambpdb -aatm -p com_solvated.top <com_solvated.crd> structure.pdb

2. Locate the atoms to be restrained e.g restraining two atoms SG of cys and C13 of your ligand.

NB: You can get the atoms and residue details of the atoms to be restrained from complex_gas.pdb as well.

3. To create the RST.dist file, you need a 7/8 column format file for the distance restraints like this;

124 ASP OD2 199 LIG N1 3.0 4.0
025 ASP 0D2 199 UNK N5 3.0 4.0

Create 8col.in file having one line from RST.dist file e.g 025 ASP 0D2 199 UNK N5 3.0 4.0

4. Use makeDIST_RST command to run the 8col.in

makeDIST_RST -ual 8col.in -pdb com_solvated.pdb -rst dist.RST

5. If (4) works properly, you should generate the dist.RST.

Repeat for each distance restrain. So, you will have dist1.RST dist2.RST and so on

You can now merge all the dist.RST files in one dist.RST file;

```
#
# 199 UNK C19 199 UNK O12 2.0 2.5
 &rst
  ixpk= 0, nxpk= 0, iat=3170,3295, r1= 1.50, r2= 2.00, r3= 2.50, r4= 3.00,
     rk2=20.0, rk3=20.0, ir6=1, ialtd=0,
 &end
#
# 124 ASH OD2 124 ASH HD2 1.0 1.2
 &rst
  ixpk= 0, nxpk= 0, iat=1986,1987, r1= 0.50, r2= 1.00, r3= 1.20, r4= 1.70,
      rk2=20.0, rk3=20.0, ir6=1, ialtd=0,
 &end
#
```

```
# 025 ASP OD2 199 UNK H82 3.0 4.0
 &rst
  ixpk= 0, nxpk= 0, iat= 417,3296, r1= 2.50, r2= 3.00, r3= 4.00, r4= 4.50,
     rk2=20.0, rk3=20.0, ir6=1, ialtd=0,
 &end
```

6. Run partial and full minimization, do heating with ibelly=1 to desired temperature in the input file.

7. MD Run looks like;

```
Production Step of MMP3 (MMMM): stage-1
 &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=1000000,
  dt=0.002,
  nmropt=1,
/
  &wt type='END' /
  DISANG=dist.RST
  DUMPAVE=chi_vs_t.170
```

8. Analyzing the results

We will start our analysis of the results by plotting temperature, total energy, kinetic energy and potential energy from the output file. Examining these properties will allow us to check that

nothing strange happened during the MD. We can check that the temperature rose smoothly and equilibrated at 300 K. Similarly changes in the kinetic energy and potential energy should be smooth and should level off at equilibrated values. Any sudden spikes in the temperature or energy plots are indicative of a problem with our simulation protocol, bad starting structure, too long a time step, inappropriate parameters etc.

In order to extract the data from the output files we can use the following perl script that automates the process (*process_mdout.perl*). We can then process the output files in one go by listing them all on the command line as follows:

chmod u+x process_mdout.perl

To analyze the data output files now copied to the analysis folder the following was then entered:

process_mdout.perl Data_01.out Data_02.out Data_03.out Data_04.out Data_05.out

Ensure you are in the directory of the Data_**.out

The next step is to locate the structure with the lowest energy, we can locate the lowest energy by looking at the summary.EPTOT file that process_mdout.perl produced. But first we need to strip off the first 50ps of the data as this represents the heating phase. Here is the file with this section removed. (*summary_EPTOT_stripped.dat.gz*). The following awk script will quickly find the lowest value in the file for us. It will print a value every time it finds a new minimum value, the last value printed will be the lowest energy found.

Move to the next step:

cat summary.EPTOT | awk '{if($2<min) {min=$2;print $1"   "min}}'

Gly the results were as follows:

the time in ns and the energy in kj cal^{-1} in the columns below respectively

2081.000   -124399.4246

2082.000   -124489.0763

2094.000   -124590.8692

2142.000   -124623.6573

2154.000   -124654.6174

2180.000   -124660.1580

2181.000   -124749.1681

So, the lowest energy is -124749.1681 kcal/mol and it occurred 2.181 ns into our simulation. We can find out where this is in our output files by using grep:
grep 2181.000 *.out using this command however only points to the binary file. To get actual output file: Gly- grep -a 2181.000 *.out
grep 324.000 *.out

Data_01.out: NSTEP = 50500 TIME(PS) = 2181.000 TEMP(K) = 300.18 PRESS = 48.9
It occurred on step 50500 of our Data_01 production step. We wrote to the mdcrd file every 500 steps (ntwx=500) (found by vi Data_01.in) during the production so this should represent frame 101 (50500/500) of data_01.mdcrd. Extract that structure using ptraj and look at it.
Simulation time= ntslim x dt= picoseconds x data files

Convert new rst file to pdb
ambpdb -p new1.top <new1.rst> new1.pdb
and view the pdb structure.

Section S3

Structure S1: The 3D ONIOM output files for the TS structures of the enzyme—substrate complexes

Sub_recog_pdb_structures.zip

References

1. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. Substrate Shape Determines Specificity of Recognition for Hiv-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes. *Structure* **2002**, *10*, 369-381.
2. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How Does a Symmetric Dimer Recognize an Asymmetric Substrate? A Substrate Complex of Hiv-1 Protease1. *Journal of molecular biology* **2000**, *301*, 1207-1220.
3. Vanderbilt University Centre for Structural Biology. *Distance Restraint between Non-Bonded Molecules*. 2008.

PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2020, **22**, 2530

Concerted hydrolysis mechanism of HIV-1 natural substrate against subtypes B and C-SA PR: insight through molecular dynamics and hybrid QM/MM studies†

Zainab K. Sanusi,[a] Monsurat M. Lawal,[a] Thavendran Govender,[b] Sooraj Baijnath,[a] Tricia Naicker,[a] Glenn E. M. Maguire,[ac] Bahareh Honarparvar *[a] and Hendrik G. Kruger *[a]

Received 16th October 2019,
Accepted 19th December 2019

DOI: 10.1039/c9cp05639d

rsc.li/pccp

It is well known that understanding the catalytic mechanism of HIV-1 PR is the rationale on which its inhibitors were developed; therefore, a better understanding of the mechanism of natural substrate hydrolysis is important. Herein, the reaction mechanism of HIV-1 natural substrates with subtypes B and common mutant in South Africa (subtype C-SA) protease were studied through transition state modelling, using a general acid–general base (GA–GB) one-step concerted process. The activation free energies of enzyme–substrate complexes were compared based on their rate of hydrolysis using a two-layered ONIOM (B3LYP/6-31++G(d,p):AMBER) method. We expanded our computational model to obtain a better understanding of the mechanism of hydrolysis as well as how the enzyme recognises or chooses the cleavage site of the scissile bonds. Using this model, a potential substrate-based inhibitor could be developed with better potency. The calculated activation energies of natural substrates in our previous study correlated well with experimental data. A similar trend was observed for the Gag and Gag-Pol natural substrates in the present work for both enzyme complexes except for the PR-RT substrate. Natural bond orbital (NBO) analysis was also applied to determine the extent of charge transfer within the QM part of both enzymes considered and the PR-RT natural substrate. The result of this study shows that the method can be utilized as a dependable computational technique to rationalize lead compounds against specific targets.

Introduction

It has been well known since the first clinical report in the 80's[1,2] that the causative agent of AIDs (acquired immunodeficiency syndrome) and its related disease is the Human immunodeficiency virus (HIV).[3,4] This HIV virus encodes an aspartic protease (Asp-PR) known as the HIV PR which promotes the replication of functional viral progeny.[5,6] HIV protease cleaves and processes the viral Gag and Gag-Pol polyprotein precursors that are required as building blocks for the final stage of the maturation process of the HIV life cycle. Blocking this action by mutagenesis or inhibitors results in the production of noninfectious, immature viral particles.[7–10] Faced with this information, potent inhibitors against the HIV PR were developed and are commercially available drugs.[10,11] Due to long-term use of the FDA approved protease inhibitors, drug resistance; polymorphisms/mutations have occurred within the PR.[12–16] It is therefore necessary to develop new potent inhibitors.

The HIV protease is a dimeric aspartic protease with two identical subunits that normally contain 99-amino acids per monomer (100 and 101-amino acids were reported by our group).[17,18] There is a conserved catalytic triad (Asp25–Thr26–Gly27) at the active site,[19,20] which usually cleaves up to nine specific substrate sites,[21] each having a different amino acids sequence in the Gag and Gag-Pol polyproteins.[5,8,22] Knowledge of the HIV PR structure and function has led the successful development of a wide range of efficient drugs designed using structure-based and substrate mechanistic methods.[23] Hence, an accurate computational transition state model is necessary to provide a better understanding of the recognition mechanism of the HIV-1 PR with regard to the cleavage of the scissile bonds. Although, there is no consensus yet on the most feasible mechanistic pathway for the HIV protease cleavage reaction, most studies have classified the mechanism into two major

[a] *Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa. E-mail: kruger@ukzn.ac.za, baha.honarparvar@gmail.com*
[b] *AnSynth PTY LTD, 498 Grove End Drive, Durban 4001, South Africa*
[c] *School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa*

† Electronic supplementary information (ESI) available: The ONIOM output files for the TS structures of the enzyme–substrate complexes are provided in PDB format and these are listed. See DOI: 10.1039/c9cp05639d

groups; the nucleophilic process and the general acid–general base (GA–GB).[22,24,25]

As described in our recent literature review,[26] the GA–GB model is the most upheld mechanism from literature for aspartic proteases, in particular the HIV PR.[27–29] This mechanism involves a catalytic water that acts as the nucleophile and attacks the carbonyl atom of the scissile peptide (C–N) bond at the active site of the Asp-PR.[25,30] The reaction proceeds with the nitrogen of the scissile bond accepting a hydrogen atom from the protonated Asp, while the unprotonated Asp ionizes the catalytic water by accepting its proton, and this process can either be a step-wise or concerted mechanism.[22,31]

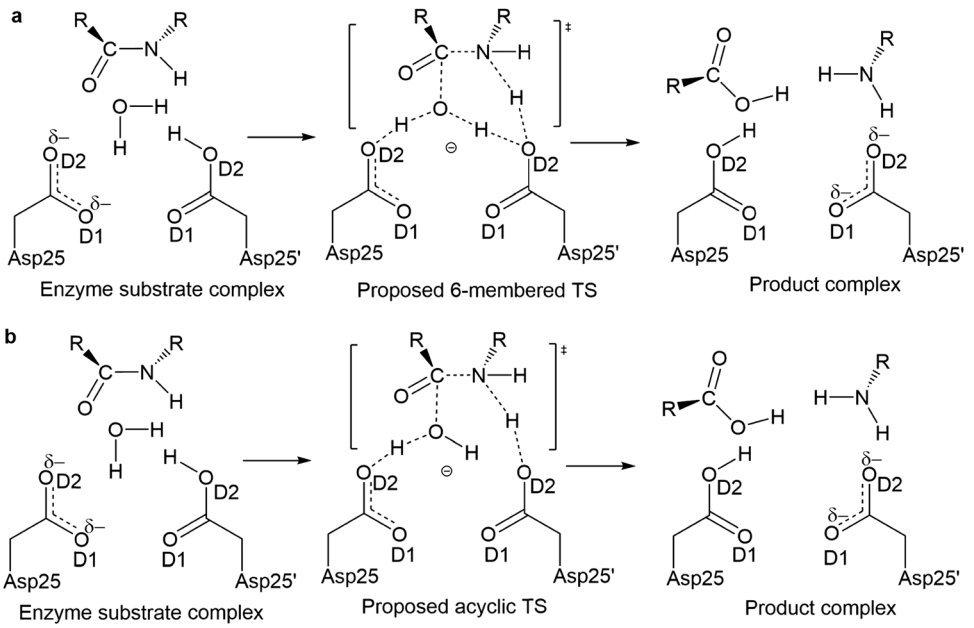
The catalytic mechanism of the enzyme–substrate and enzyme–intermediate complexes of HIV PR have been theoretically studied using molecular mechanics (MM) and molecular dynamics (MD) modelling.[32–34] The application of QM/MM methods in studying the HIV-1 PR mechanism have entailed the use of hybrid Car–Parrinello (QM) and MM,[35,36] classical QM/MM MD,[9,24] enhanced QM/MM MD umbrella sampling,[30] QM/MM interface algorithm[37] and ONIOM QM/MM approach.[38,39]

The reactivity of the HIV-1 PR model was examined in 2011 by Garrec *et al.*, to specifically validate the accuracy of the latest density functional theory (DFT) methods.[40] DFT method has been considered suitable to elucidate mechanistic pathways and obtain results closer to estimated experimental values,[40–43] and the combination of B3LYP[44,45]/6-31++G(d,p)[46,47] or B3LYP/6-311++G(d,p) was implemented. The hybrid approach of Our own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM)[48,49] method using the DFT functional has been used to study the inhibition mechanisms of matrix metalloproteinase 2,[50] L,D transpeptidase[51,52] and recently, in our group, the concerted catalytic mechanism of HIV PR.[38,39] This multi-layered approach allows treatment of large molecular systems in different research areas. Herein, the catalytic active site was treated at a high DFT level of theory, while the rest of the system is treated with MM (AMBER)[53] level of theory.

Literature support exits for a concerted hydrolysis reaction mechanism for similar chemical reactions[41,43,51,54–56] and enzymatic reactions. We previously[39] calculated theoretical transition structures involving a 6-membered ring and an acyclic model in a one-step concerted reaction process for HIV-1 PR (Scheme 1). The activation free energies for hydrolysis of the Abz-Arg-Val-Nle-Phe(NO_2)-Glu-Ala-Nle-NH_2 substrate (which mimics the CA/p2 natural substrate) by both subtypes B/C-SA HIV-1 PR was determined experimentally.[18,57–59] In which, cyclic transition states (TSs)[25,30] with activation free energies ($\Delta G^{\ddagger}$) of between 30[25] and 43.5[30] kcal mol^{-1} were reported, showing higher values comparable to the experimental data (14.9–21.0 kcal mol^{-1} depending on the peptide sequence).[27,60–65] It should also be noted that two theoretical models[25,30] were found in literature for the HIV-1 PR cleavage of peptide with respect to the one-step concerted mechanism.

In our earlier studies,[38,39] the one-step concerted acyclic reaction mechanism of HIV-1 PR (Scheme 1b) provided an optimal basis for substrate recognition and cleavage. From our most recent work where only two substrates (MA–CA and RH–IN) were considered with subtypes B/C-SA, respectively, $\Delta G^{\ddagger}$ values of 15.2/13.5 kcal mol^{-1} and 19.7/18.3 kcal mol^{-1} were obtained, respectively.[38] It was also realized that a longer natural substrate sequence (P5–P5′) was required for favorable recognition; hence a 10-amino peptide sequence was used in the present investigation.

Herein, seven of the remaining Gag and Gag-Pol natural substrate polyproteins are complexed with two HIV-1 enzymes;



Scheme 1 The one-step concerted enzymatic mechanism of HIV-1 PR with an idealised amide precursor.[38,39] (a) Proposed concerted 6-membered cyclic HIV-1 PR enzymatic mechanism. (b) Proposed acyclic HIV-1 PR enzymatic mechanism.

Table 1 Recognition sequences for natural substrates cleaved by the HIV-1 protease[67,68]

Peptide sequences cleavage domain	Cleavage domain
Cleavage sites in Gag	
P5 P4 P3 P2 P1 * P1′ P2′ P3′ P4′ P5′	
Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-Asn	**MA-CA**
Lys-Ala-Arg-Val-Leu*Ala-Glu-Ala-Met-Ser	CA-p2
Pro-Ala-Thr-Ile-Met*Met-Gln-Arg-Gly-Asn	p2-NC
Glu-Arg-Gln-Ala-Asn*Phe-Leu-Gly-Lys-Ile	NC-p1
Arg-Pro-Gly-Asn-Phe*Leu-Gln-Ser-Arg-Pro	p1–p6
Cleavage sites in Gag-Pol	
Val-Ser-Phe-Asn-Phe*Pro-Gln-Ile-Thr-Leu	TF-PR
Cys-Thr-Leu-Asn-Phe*Pro-Ile-Ser-Pro-Ile	PR-RT
Gly-Ala-Glu-Thr-Phe*Tyr-Val-Asp-Gly-Ala	RT-RH
Ile-Arg-Lys-Ile-Leu*Phe-Leu-Asp-Gly-Ile	**RH-IN**

The asterisk (*) denotes the scissile bond. Matrix-capsid; MA-CA, capsid-p2; CA-p2, p2-nucleopsid; p2-NC, nucleopsid-p1; NC-p1, *trans* frame peptide-protease; TF-PR, protease-reverse transcriptase; PR-RT, reverse transcriptase-RNAseH; RT-RH, RNAseH-integrase; RH-IN. The P1, P2…P*n* to P1′, P2′…P*n*′ represents the amino acid residues of the substrate and are counted from the cleavage point. *N. B.*: the activation free energies of the substrates in bold has been determined in our previous paper.[38]

subtype B[11] and a common mutant in South Africa (subtype C-SA).[66] This study was designed to expand our understanding of the reaction mechanism of the HIV-1 PR, since its mutants as well as the active enzyme recognize and cleave the same polypeptide precursor. The HIV-1 PR cleavage sites in the Gag and Gag-Pol polyproteins are represented in Fig. S1 (ESI†) and Table 1.

HIV-1 protease hydrolysis of the natural substrate involves cleavage and formation of chemical bonds;[25] hence we employed the use of ONIOM,[48,69,70] a hybrid QM/MM approach for this study. The catalytic process follows a concerted, general acid–general base mechanism, with a mono-protonation of the Asp at the active site, the same as the model adopted in our previous work.[17,71,72] Also, to provide more in-depth knowledge of the enzyme–substrate reactivity, the natural bond orbital[73] was used to identify the charge transfer between the QM region of the enzyme proteases (subtype B and C-SA) and PR-RT natural substrate.

In this study, the relative thermodynamics and kinetics of the concerted acyclic cleavage mechanism of the natural substrates by HIV-1 subtypes B and C-SA PRs was studied using ONIOM (B3LYP/6-31++G(d,p):AMBER) model and energetics are reported in kcal mol^{-1}.

Computational methodology

The X-ray structural coordinates of the natural substrates were taken from the Protein Data Bank (PDB) with the following codes; 1KJF (p1–p6),[68] 1KJ7 (p2-NC),[68] 1KJG (RT-RH),[68] 1F7A (CA-p2),[67] 1TSU (NC-p1),[74] 3OU4 (TF-PR),[75] 3OU3 (PR-RT).[75] The initial structural models for subtype B and C-SA HIV-1 PR were taken at 1.8 and 2.7 Å resolution from PDB bank (codes: 2P3B[11] and 3U71[66]) respectively. The 3D structures for all the selected enzyme–substrates complexes were constructed by superimposing the available crystal structure of the natural substrates with subtype B (details provided in the ESI†) and C-SA PR, using the same overlay method described in our previous papers.[17,72] Since the involvement of water is an important aspect for the HIV-1 PR proteolysis, a catalytic water molecule (PDB code: 1LV1[76]) closest to the active site is also superimposed on the generated enzyme–substrate complexes using PyMOL.[77] PyMOL measures the root mean square (RMS), which is helpful to evaluate how well the enzyme–substrate complexes were overlaid, and an optimal superimposition is considered satisfactory if the RMS is less than 2 Å.[78–80] The RMS values of the superimposed enzyme–substrate complexes were in the range of 1.5–1.7 Å for all 14 systems (7 substrates with two HIV strains). In this strategy, the substrate maintains the same pose in the active binding sites of the complexes and serve as a consistent starting structure for all complexes. Thus, the structures of all enzyme–substrate complexes were refined by removing the ions and crystallographic waters from the complexes except for the catalytic water in the active site.

An area of contention in modelling of HIV PR is the protonation state of the catalytic aspartate (Asp25/25′) at the active site and this has been extensively discussed in literature.[40,81,82] Both *in vivo* and *in vitro* studies show that the catalytic process of HIV PR functions at a pH of 2–7.[26,81,83,84] Hence, several different protonation states at the active site have been proposed in literature.[26] Using PROPKA[85] in assigning the protonation state for the catalytic Asp, a mono-protonation model at Asp25′ was adopted and this approach is supported by the majority of studies.[25,27,71,84,86,87]

The empirical PROPKA server[85] was used to calculate the pK_a values of the titratable amino acids in assigning protonation to both enzymes. A mono-protonated catalytic Asp25/25′ was induced in the active site at a physiological pH 7, using GaussView.[88] It is notable that the standard pK_a values of ionizable groups can be shifted due to protein environments.[89] The proton (H^+) of the protonated Asp in the enzyme–substrate complexes is positioned midway to facilitate its donation to the nitrogen of the scissile bond, this position has been measured to be energetically stable *via* molecular dynamics (MD) simulations.[34]

The MD simulations follows a similar protocol as described by Ribeiro *et al.*,[90] and explained in our previous paper.[38] It should be noted that MD simulations was carried out to explore all the possible conformational space to obtain a stable starting structure for subsequent ONIOM TS calculations. The initial stage involves minimizations at 2500 cycles which was followed by the heating stage at an increasing temperature from 0 K to 600 K. Constraint was applied to counteract the SHAKE failure that might occur because of the increase in temperature at the heating steps. It has been established before that higher temperatures are necessary to overcome energy barriers of constrained molecules.[87] This is required to explore all possible enzyme–substrate conformations, the constraint allows only the enzyme to be frozen and the substrate (10 amino acid sequence) was subjected to heating for up to 600 K. Visual inspections of the substrate during the MD simulation shows that an extensive enzyme–substrate interactions were achieved. As was done in our previous work,[38] a 3-stage stepwise equilibration was performed at 10, 5 and 2 ns, and the constraint forces was reduced at each step. While at the production stage,

a constraint distance of 3–4 Å was applied between the OH of the water (O_{wat}) with carbonyl of the scissile bond (C_{sub}) and the OH of the protonated Asp (OD2 Asp25′) with the nitrogen of the scissile bond (N_{sub}) was fixed and run for 10 ns. An unconstrained MD calculation at 300 K was later performed for 2 ns and a representative snapshot of the complex with lowest energy conformation was attained. Subsequently, the lowest energy conformation for all complexes were overlaid with the enzymes (B and C-SA PR respectively) that did not undergo constraint MD and was taken for further ONIOM calculations for an unconstrained acyclic concerted TS optimization. TS starting structures were generated constraining the interatomic distances where bond breaking/formation takes place, as described before.[38,91]

Due to the different trend found in the theoretical data *versus* experimental results for two of the substrates (PR-RT and TF-PR), triplicate MD[92] runs were also performed. This was done by (a) varying the random seed to create a new starting MD simulation with a different atomic velocity of 10 Å water box, (b) then creating another starting MD simulation with different atomic coordinates as well as different atomic velocities and changing the size of solvent/water box to 12 Å. As aforementioned, an unconstrained MD calculation at 300 K was performed for 2 ns for all the three cases and a representative snapshot of the complexes with the lowest energy was attained. These were used for subsequent ONIOM transition state calculations.

The active Asp25/25′ residues, catalytic water, and all the substrates used were considered at a high quantum mechanics (QM) layer; (DFT/B3LYP[44,45]/6-31++G(d,p)[46,47]), while the rest of the system were considered at a molecular mechanics (MM) layer; (AMBER[53]) for the ONIOM calculations (Fig. 1). The electric field is taken into consideration as the reaction centre involves the transfer of protons due to transition state modelling, therefore an ONIOM electronic embedding conformation was specified during the calculations.[70]

ONIOM free energy calculations

The inclusion of explicit electrons in energy calculations is necessary to model the breaking and formation of chemical bonds,[21] hence a multi-layered method is adopted for the entire enzyme.[9] A two-layer ONIOM[48,93] method was applied to calculate the activation free energies of the catalysed reaction by defining the systems into two regions, one classical and the other quantum mechanical.[86,94,95] Previous studies have shown that the hybrid QM/MM methods are adequate for the HIV-1 PR catalytic mechanism and the accuracy of the DFT/B3LYP level of theory has been tested for HIV-1 PR reactivity.[9,94–96]

The enzyme–substrate complex structures for obtaining starting optimized reactant states are prepared from energy minimization calculations using the Amber forcefield,[97] this is necessary to allow both the substrates and catalytic water to adjust to the active binding site. The lowest energy conformation taken from the constrained MD that was initially performed for all complexes were superimposed with an unconstrained subtype B and C-SA HIV-1 PR and this was further used for ONIOM calculations to obtain an unconstrained acyclic concerted TS optimization. A full unconstrained geometry optimizations of the input structure complexes in gas phase using ONIOM (B3LYP/6-31++G(d,p):AMBER) model were carried out, vibrational frequencies[98] for the different complexes were computed to assess the TS structures with one negative eigenvalue. The TS models were determined by having only one negative frequency, the atomic vibrations associated with the negative eigenvalue were visually inspected to ensure the expected bond cleavage and formation within the enzyme–substrate complex. GaussView 5.0.8[88] was used for both pre and post-processor graphic interface analyses in this study, and the Gaussian 16 package[99] was used for all calculations.

The chemical properties such as thermodynamics, that comprises of energy, entropy contribution and enthalpy can be estimated computationally.[100] Herein, the activation free energy, enthalpy and temperature-dependent change in entropy (T = 298.15 K) is represented as; $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $T\Delta S^{\ddagger}$, respectively. The following equation gives the activation free energy attained from frequency calculations for the natural substrate polypeptide sequences:

$$\Delta G^{\ddagger}_{\text{substrate}} = \Delta G_{\text{complex}} - \Delta G_{\text{protease}} - \Delta G_{\text{substrate}} \quad (1)$$

Natural bond orbital (NBO) analysis

The natural bond orbital[101] offers a robust mathematical method for investigating charge transfer that occur between hydrogen bond atoms through a set of occupied Lewis and unoccupied non-Lewis localized structural orbitals.[102–104] The second-order perturbation between the proton donor (i) and

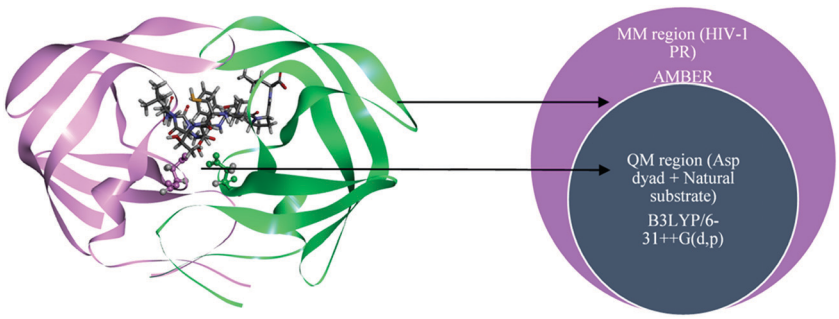


Fig. 1 Schematic representation of the applied two-layered ONIOM model (B3LYP/6-31++G(d,p):AMBER) for HIV-1 PR-natural substrate (P5–P5′) complex (acyclic TS with one catalytic water molecule). The model includes 77 atoms at the QM region and 1512 atoms at the MM region. All the 3D structures of the enzyme–substrate complexes considered are provided with the ESI† (Gaussian input file format).

proton receptor (j) is used to predict the stabilization energy also known as the delocalization strength E^2 and this is measured by eqn (2)[105,106] as:

$$E^{(2)} = \Delta E_{ij} = q_i \frac{F(i,j)^2}{e_j - e_i} \quad (2)$$

where q_i is the occupancy number of the donor orbital, e_j and e_i are orbital energies (diagonal elements) and $F(i,j)^2$ is the off-diagonal elements of Fock matrix.

Results and discussion

Our previous protocol of transition state modeling[38] on the inhibition mechanism of HIV PR and the activation energies of the scissile *versus* non-scissile bonds was utilized to demonstrate the recognition pattern of different natural substrates against HIV-1 subtypes B and C-SA. The HIV-1 PR cleaves at nine different sites of the Gag and Gag-Pol protein leading to different amino acid sequence. Earlier, the MA-CA and RH-IN domain have been investigated[38] and it was discovered that the HIV-1 PR specifically cleaves at the scissile region of its natural substrates. This prompted us to investigate all the remaining natural substrates of PR and determine the rate of their hydrolysis and cleavage process using the P5–P5′ sequence. A concerted general acid–base mechanism as described in our previous work[38,39] was applied, where the protonated catalytic Asp25′ acts as the acid and the unprotonated Asp25 acts as the base in a one-step acyclic reaction process (Scheme 1b). The catalytic function of the Asp25/25′ groups as well as the nucleophilic attack of the water molecule on the scissile (C–N) bond and other amide bonds occurred simultaneously. The concerted acyclic reaction proceeds with an attack from the catalytic water (OH) on the carbonyl atom of the scissile bond to form COOH, meanwhile Asp25′ donates a proton to the nitrogen atom of the scissile bond which then becomes a base. Subsequently Asp25 accepts a proton from the nucleophilic water yielding OH to become acidic.

The free energy values were computed for the enzyme–substrate complexes and are ranked in terms of their activation free energies for the Gag and Gag-Pol polyproteins for subtypes B and C-SA HIV-1 PR (Tables 2 and 3). The calculated activation energies for MA-CA and RH-IN substrates from our previous paper[38] were included to get an overall view. The theoretical results (Table 2) are in reasonable agreement with the available experimental results (within experimental error). Note that the standard deviation of the theoretical activation energies for PR-RT and TF-PR were as high as 1.0 and 0.7 kcal mol^{-1} respectively (see ESI,† Table S1).

The computed results for the Gag polyproteins (Table 2) have the MA-CA natural substrate with the smallest activation energy ($\Delta G^{\ddagger}$ of 15.2 kcal mol^{-1}), and that of NC-p1 the largest (20.0 kcal mol^{-1}). This largest $\Delta G^{\ddagger}$ for NC-p1 is expected based on the study carried out by Moses *et al.*,[91] where the NC-p1 natural substrate processing is not only the slowest, but also the rate-limiting cleavage step in the maturation of the Gag polyprotein.[91]

Table 2 Relative thermodynamics and kinetic parameters for the one-step concerted acyclic TSs involved in the cleavage mechanism of the P5–P5′ natural substrates for scissile bonds by subtype B HIV-1 PR using ONIOM [B3LYP/6-31++G(d,p)]

HIV-1 PR B	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta G^{a\ddagger}$
Gag					
MA-CA	**−28.6**	**−23.5**	**−38.7**	**15.2**	**15.7**
CA-p2	−30.0	−31.0	−47.7	16.8	17.0
p2-NC	−22.5	−24.0	−43.1	19.2	—
p1–p6	−23.8	−24.6	−44.3	19.7	17.6
NC-p1	−19.6	−20.3	−40.4	20.0	18.7
Gag-Pol					
RH-IN	**−9.4**	**−10.6**	**−30.3**	**19.7**	**18.42**
RT-RH	−23.3	−24.3	−39.1	14.7	—
TF-PR	−29.0	−32.1	−48.6	16.5	15.4
PR-RT	−36.0	−39.0	−53.4	14.5	15.6

Values are reported in kcal mol^{-1} relative to the sum of separated reactants. $\Delta E^{\ddagger}$ = total energy, $\Delta H^{\ddagger}$ = enthalpy, $T\Delta S^{\ddagger}$ = entropy change in temperature, $\Delta G^{\ddagger}$ = calculated activation free energy, $\Delta G^{a\ddagger}$ = experimental activation free energies taken from literature for subtype B HIV-1 PR.[62,64] All the 3D structures of the enzyme–substrate complexes considered are provided with the ESI (Gaussian input file format). *N. B.*: the activation free energies of the substrates in bold has been determined in our previous paper.[38]

Cleavage of the scissile bonds of the aforementioned substrates by both subtype B and C-SA HIV-1 PR gave a significant entropy variation as seen from the results of the TS structures (Tables 2 and 3). It should be noted that more negative entropies (higher entropy penalty) is an indication of greater restriction of movement for the substrate in the active site.[72,107] This could be attributed to steric restrictions and strong non-covalent interactions for certain parts of the substrate.[77,104]

Even though the scissile bond of both substrates (PR-RT and TF-PR) consists of hydrophobic amino acids (Phe*Pro), for TF-PR

Table 3 Relative thermodynamics and kinetic parameters for the one-step concerted acyclic TSs involved in the cleavage mechanism of the P5–P5 natural substrates for scissile bonds by subtype C-SA HIV-1 PR using ONIOM [B3LYP/6-31++G(d,p)]

HIV-1 PR C-SA	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$	$\Delta G^{\ddagger}$	$\Delta G^{a\ddagger}$
Gag					
MA-CA	**−40.2**	**−39.3**	**−52.8**	**13.5**	**15.7**
CA-p2	−42.2	−49.1	−65.9	16.8	17.0
p2-NC	−45.3	−53.4	−74.0	20.6	—
p1–p6	−45.0	−53.6	−74.6	21.0	17.6
NC-p1	−26.1	−29.5	−52.1	22.6	18.7
Gag-Pol					
RH-IN	**−25.6**	**−29.3**	**−47.7**	**18.3**	**18.4**
RT-RH	−34.6	−39.1	−53.4	14.8	—
TF-PR	−39.1	−44.9	−60.7	15.8	15.4
PR-RT	−44.1	−49.7	−66.1	16.4	15.6

Values are reported in kcal mol^{-1} relative to the sum of separated reactants. $\Delta E^{\ddagger}$ = total energy, $\Delta H^{\ddagger}$ = enthalpy, $T\Delta S^{\ddagger}$ = entropy change in temperature, $\Delta G^{\ddagger}$ = activation free energy, $\Delta G^{a\ddagger}$ = experimental activation free energies taken from literature for subtype B HIV-1 PR.[62,64] All the 3D structures of the enzyme–substrate complexes considered are provided with the ESI (Gaussian input file format). *N. B.*: the activation free energies of the substrates in bold has been determined in our previous paper.[38]

(16.51 kcal mol^{-1}), a hydrophilic amino acids Asn and Gln is found at the P2–P2′ position of the scissile bond, which may slow down the hydrolysis. While, for PR-RT a polar amino acid (Asn) at P2 precedes the phenylalanine (P1 position) and the Pro (P1′ position) is followed by a hydrophobic amino acid (Ile) at P2′. Although, the PR-RT and TF-PR complexes gave a reversed order for the theoretical activation energies (compared to the experimental results), the results are within the experimental error.

While both subtype B and C-SA HIV-1 PR recognize and cleave these natural substrates, the rate of proteolysis of the enzymes are different as can be seen from their activation free energies. This could be because of the eight-point mutations in C-SA compared to the subtype B (Tables 2 and 3). The theoretical results (Table 2) for the activation energies of the natural substrates with subtype C-SA are also in reasonable agreement with the available experimental results (within experimental error).

Similar to the subtype B–substrate complexes, the MA-CA and NC-p1 (Gag polyproteins) complexed with the subtype C-SA HIV-1 PR also gave the smallest and largest activation energies ($\Delta G^{\ddagger}$ of 13.5 and 22.6 kcal mol^{-1}, respectively). The computed data for the Gag polyproteins of subtype C-SA follow the same trend than for subtype B HIV-1 PR.

While the same trend was observed for both HIV-1 PR subtype B and C-SA–substrate complexes, an outlier was observed for the Gag-Pol substrates. A reduced activation energy was found for PR-RT in the subtype B complex (14.5 kcal mol^{-1}) compared to a much-increased activation energy in the C-SA complex (16.4 kcal mol^{-1}). Hence, NBO analysis was carried out on the PR-RT substrate to further understand the electronic factors contributing to the reactivity of the both proteases.

Natural bond orbital analysis

To further explore the hydrogen bond stabilization energies of both subtype B and C-SA complexed with PR-RT natural substrate, NBO analysis was performed at the DFT B3LYP/6-31++G(d,p) for the QM region and this is reported in Table 4. The hydrogen bond formation in the studied complexes indicates that certain transfer of electronic charge occurs during the transition state modelling (Fig. 2). The intensive interactions of the hydrogen bonds are directly proportional to the second-order perturbation (E^2) value, hence the higher the E^2 value the greater charge transfer and consequently more stable the interaction. The charge transfers particularly from lone pair (LP) to unoccupied lone pair orbital (LP*) of the subtype B–PR-RT complex gave a higher E^2 value, which provides stability to the transition state (TS) model structure with a reduced activation energy of 14.5 kcal mol^{-1} (Table 2).

This clearly indicates that intermolecular interactions in the concerted acyclic mechanism of the PR-RT substrate in the subtype B enzyme are greater compared to the intermolecular interactions of the PR-RT substrate in the subtype C-SA PR. These results explain the anomaly for the theoretical results that we have obtained.

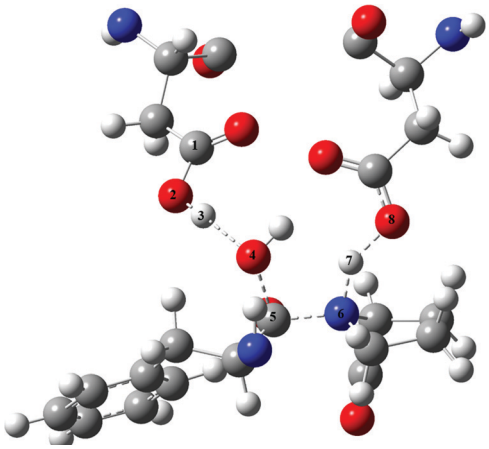


Fig. 2 Molecular plot of the interaction in the QM region (catalytic Asp25/25′ with PR-RT natural substrate) conforming to the most important charge transfer in the reaction coordinate.

Table 4 Second-order perturbation E^2 in kcal mol^{-1} conforming to the most important charge transfer in the reaction coordinate (donor → acceptor) at the B3LYP/6-31++G(d,p) level of theory in the QM region of the PR-RT enzyme complexes

Complex	Donor NBO	Acceptor NBO	E^2 (kcal mol^{-1})
Subtype B–PR-RT	LP(O_8)	LP*(H_7)	122.4
	LP(N_6)	LP*(H_7)	213.6
	LP*(H_7)	BD*(C_9–O_8)	2.5
Subtype C–PR-RT	LP(O_8)	LP*(H_7)	14.3
	LP(N_6)	LP*(H_7)	181.0
	LP*(H_7)	BD*(C_9–O_8)	2.4

LP = lone pair, LP* = unoccupied lone pair, BD* = anti-bonding orbital.

Conclusion

This study employs a two-layered ONIOM (B3LYP/6-31G++(d,p):AMBER) model to calculate the TS energies *via* a concerted mechanism of hydrolysis. It examines the subtype B and C-SA HIV-1 PR recognize their scissile bonds. Based on experimental results and our previous paper a longer peptide is required to recognize the substrate effectively and thus obtain a more accurate model for hydrolysis. Therefore, a P5–P5′ amino acid sequence was used to compute the free energy profiles of the complexes. The reaction mechanism follows a concerted acyclic TS model leading to product and activation free energies that are in close proximity with experimental values for hydrolysis by HIV-1 PR. It is evident from the computed results that both proteases complexed with the natural substrate follow a similar activation energy trend at the Gag and Gag-Pol. The exception is with the PR-RT substrate, the extent of charge transfer between the natural substrate (PR-RT) and the active residues in close contact with the substrate (derived by second-order perturbation theory of NBO analysis) supports the observation of the varying activation energies of the PR-RT substrate against both HIV-1 PR subtypes. While the CA-p2 substrate gave a lower calculated $\Delta G^{\ddagger}$ than the experimental results, RT-RH substrate had a lower calculated value of $\Delta G^{\ddagger}$ at the Gag-Pol for both subtypes considered. The NC-p1 substrate gave an increase $\Delta G^{\ddagger}$ (at Gag and Gag-Pol) for both PR subtypes compared to the experimental results. The catalytic

mechanism of HIV-1 PR is the rationale on which its inhibitors were developed. It is expected that an increase in the activation energy of the natural substrate will retard the rate of hydrolysis and thus serve as an improved model for a competitive inhibitor. Thus, it is important to understand which of the HIV-1 natural substrates best binds to both subtype B and C-SA PRs in finding potential lead compounds, necessary for drug discovery. We therefore, propose that future potent inhibitors could be developed to mimic the NC-p1 substrate for both subtype B and C-SA HIV-1 PR based on its mechanism of reaction. This could be related to not only the slow rate of its hydrolysis (NC-p1) but also its rate-limiting cleavage step in the maturation of the Gag, as seen from the compiled results and related experimental data. This model can also potentially provide information on how electronic and steric effects will change the activation energies and the reaction rate of natural substrates.

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgements

The authors thank College of Health Sciences, University of KwaZulu-Natal, Asphen Pharmacare, Medical Research Council and the National Research Foundation (all in South Africa) for financial support. We are also grateful to the Centre for High Performance Computing (www.chpc.ac.za) for computational resources.

References

1 D. Klatzmann, E. Champagne, S. Chamaret, J. Gruest, D. Guetard and T. Hercend, *et al.*, T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV, *Nature*, 1983, **312**(5996), 767–768.
2 R. C. Gallo, P. S. Sarin, E. Gelmann, M. Robert-Guroff, E. Richardson and V. Kalyanaraman, *et al.*, Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS), *Science*, 1983, **220**(4599), 865–867.
3 F. Barré-Sinoussi, J.-C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret and J. Gruest, *et al.*, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science*, 1983, **220**(4599), 868–871.
4 K.-C. Chou, A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins, *J. Biol. Chem.*, 1993, **268**(23), 16938–16948.
5 N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach and R. Dixon, *et al.*, Active human immunodeficiency virus protease is required for viral infectivity, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**(13), 4686–4690.
6 M. A. Navia, P. M. Fitzgerald, B. M. McKeever, C.-T. Leu, J. C. Heimbach and W. K. Herber, *et al.*, Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature*, 1989, **337**(6208), 615.
7 A. Wlodawer and J. W. Erickson, Structure-based inhibitors of HIV-1 protease, *Annu. Rev. Biochem.*, 1993, **62**(1), 543–585.
8 I. T. Weber, D. S. Cavanaugh and R. W. Harrison, Models of HIV-1 protease with peptides representing its natural substrates, *THEOCHEM*, 1998, **423**(1-2), 1–12.
9 D. C. Chatfield, K. P. Eurenius and B. R. Brooks, HIV-1 protease cleavage mechanism: a theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential, *THEOCHEM*, 1998, **423**(1), 79–92.
10 N. A. Roberts, J. A. Martin, D. Kinchington, A. V. Broadhurst, J. C. Craig and I. B. Duncan, *et al.*, Rational design of peptide-based HIV proteinase inhibitors, *Science*, 1990, **248**(4953), 358–361.
11 D. J. Kempf, D. W. Norbeck, L. Codacovi, X. C. Wang, W. E. Kohlbrenner and N. E. Wideburg, *et al.*, Structure-based, C2 symmetric inhibitors of HIV protease, *J. Med. Chem.*, 1990, **33**(10), 2687–2689.
12 J. L. Martinez-Cajas and M. A. Wainberg, Protease inhibitor resistance in HIV-infected patients: molecular and clinical perspectives, *Antiviral Res.*, 2007, **76**(3), 203–221.
13 J. P. Vacca and J. H. Condra, Clinically effective HIV-1 protease inhibitors, *Drug Discovery Today*, 1997, **2**(7), 261–272.
14 M. Witvrouw, C. Pannecouque, W. M. Switzer, T. M. Folks, E. De Clercq and W. Heneine, Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis, *Antiviral Ther.*, 2004, **9**(1), 57–66.
15 B. Rodés, C. Toro, J. A. Sheldon, V. Jiménez, K. Mansinho and V. Soriano, High rate of proV47A selection in HIV-2 patients failing lopinavir-based HAART, *AIDS*, 2006, **20**(1), 127–129.
16 M. Ntemgwa, B. G. Brenner, M. Oliveira, D. Moisi and M. A. Wainberg, Natural polymorphisms in the human immunodeficiency virus type 2 protease can accelerate time to development of resistance to protease inhibitors, *Antimicrob. Agents Chemother.*, 2007, **51**(2), 604–610.
17 Z. K. Sanusi, T. Govender, G. E. Maguire, S. B. Maseko, J. Lin and H. G. Kruger, *et al.*, An insight to the molecular interactions of the FDA approved HIV PR drugs against L38L↑ N↑ L PR mutant, *J. Comput.-Aided Mol. Des.*, 2018, **32**(3), 459–471.
18 S. B. Maseko, S. Natarajan, V. Sharma, N. Bhattacharyya, T. Govender and Y. Sayed, *et al.*, Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies, *Protein Expression Purif.*, 2016, **122**, 90–96.
19 F. D. Veronese, A. L. DeVico, T. D. Copeland, S. Oroszlan, R. C. Gallo and M. Sarngadharan, Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene, *Science*, 1985, **229**(4720), 1402–1405.
20 R. V. Stanton, D. S. Hartsough and K. M. Merz Jr, Calculation of solvation free energies using a density functional/molecular dynamics coupled potential, *J. Phys. Chem.*, 1993, **97**(46), 11868–11870.
21 S. C. Pettit, S. F. Michael and R. Swanstrom, The specificity of the HIV-1 protease, *Perspect. Drug Discovery Des.*, 1993, **1**(1), 69–83.

22 J. Trylska, P. Grochowski and J. A. McCammon, The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease, *Protein Sci.*, 2004, **13**(2), 513–528.
23 J. W. Erickson and S. K. Burt, Structural mechanisms of HIV drug resistance, *Annu. Rev. Pharmacol. Toxicol.*, 1996, **36**(1), 545–571.
24 D. C. Chatfield and B. R. Brooks, HIV-1 protease cleavage mechanism elucidated with molecular dynamics simulation, *J. Am. Chem. Soc.*, 1995, **117**(20), 5561–5572.
25 H. Park, J. Suh and S. Lee, Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism, *J. Am. Chem. Soc.*, 2000, **122**(16), 3901–3908.
26 M. Lawal, Z. Sanusi, T. Govender, G. E. Maguire, B. Honarparvar and H. Kruger, From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets, *Curr. Med. Chem.*, 2018, **26**, 1–34.
27 L. J. Hyland, T. A. Tomaszek Jr and T. D. Meek, Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism, *Biochemistry*, 1991, **30**(34), 8454–8463.
28 V. Antonov, L. Ginodman, Y. V. Kapitannikov, T. Barshevskaya, A. Gurova and L. Rumsh, Mechanism of pepsin catalysis: general base catalysis by the active-site carboxylate ion, *FEBS Lett.*, 1978, **88**(1), 87–90.
29 J. S. Fruton, The mechanism of the catalytic action of pepsin and related acid proteinases, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1976, **44**, 1–36.
30 A. Krzemińska, V. Moliner and K. Świderek, Dynamic and electrostatic effects on the reaction catalyzed by HIV-1 protease, *J. Am. Chem. Soc.*, 2016, **138**(50), 16283–16298.
31 J. Trylska, P. Bała, M. Geller and P. Grochowski, Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease, *Biophys. J.*, 2002, **83**(2), 794–807.
32 A. Tropsha and J. Hermans, Application of free energy simulations to the binding of a transition-state-analogue inhibitor to HTV protease, *Protein Eng., Des. Sel.*, 1992, **5**(1), 29–33.
33 A. Goldblum, M. Glick and A. Rayan, Determining proton positions in an enzyme-inhibitor complex is a first step for theoretical mechanistic studies of aspartic proteinases, *Theor. Chim. Acta*, 1993, **85**(1–3), 231–247.
34 R. W. Harrison and I. T. Weber, Molecular dynamics simulations of HIV-1 protease with peptide substrate, *Protein Eng., Des. Sel.*, 1994, **7**(11), 1353–1363.
35 A. Laio, J. VandeVondele and U. Rothlisberger, A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations, *J. Chem. Phys.*, 2002, **116**(16), 6941–6947.
36 S. Piana, D. Bucher, P. Carloni and U. Rothlisberger, Reaction mechanism of HIV-1 protease by hybrid car-parrinello/classical MD simulations, *J. Phys. Chem. B*, 2004, **108**(30), 11139–11149.
37 P. Altoè, M. Stenta, A. Bottoni and M. Garavelli, A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction, *Theor. Chem. Acc.*, 2007, **118**(1), 219–240.
38 Z. Sanusi, M. Lawal, T. Govender, G. Maguire, B. Honarparvar and H. Kruger, Theoretical model for HIV-1 PR that accounts for substrate recognition and preferential cleavage of natural substrates, *J. Phys. Chem. B*, 2019, **123**, 6389–6400.
39 M. M. Lawal, Z. K. Sanusi, T. Govender, G. F. Tolufashe, G. E. Maguire and B. Honarparvar, *et al.*, Unraveling the concerted catalytic mechanism of the human immunodeficiency virus type 1 (HIV-1) protease: a hybrid QM/MM study, *Struct. Chem.*, 2018, **30**, 409–417.
40 J. Garrec, P. Sautet and P. Fleurat-Lessard, Understanding the HIV-1 protease reactivity with DFT: what do we gain from recent functionals?, *J. Phys. Chem. B*, 2011, **115**(26), 8545–8558.
41 M. M. Lawal, T. Govender, G. E. Maguire, B. Honarparvar and H. G. Kruger, Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study, *J. Mol. Model.*, 2016, **22**(10), 235–246.
42 M. M. Lawal, T. Govender, G. E. Maguire, H. G. Kruger and B. Honarparvar, DFT study of the acid-catalyzed esterification reaction mechanism of methanol with carboxylic acid and its halide derivatives, *Int. J. Quantum Chem.*, 2018, **118**(4), 25497–25508.
43 Z. Fakhar, T. Govender, G. Lamichhane, G. E. Maguire, H. G. Kruger and B. Honarparvar, Computational model for the acylation step of the β-lactam ring: potential application for l, d-transpeptidase 2 in mycobacterium tuberculosis, *J. Mol. Struct.*, 2017, **1128**, 94–102.
44 A. D. Becke, Density-functional thermochemistry III. The role of exact exchange, *J. Chem. Phys.*, 1993, **98**(7), 5648–5652.
45 C. Lee, W. Yang and R. G. Parr, Development of the colle-salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**(2), 785–789.
46 V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, 6-31G* basis set for atoms K through Zn, *J. Chem. Phys.*, 1998, **109**(4), 1223–1229.
47 V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern and L. A. Curtiss, 6-31G* basis set for third-row atoms, *J. Comput. Chem.*, 2001, **22**(9), 976–984.
48 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives, *THEOCHEM*, 1999, **461**, 1–21.
49 T. Vreven and K. Morokuma, On the application of the IMOMO (integrated molecular orbital+ molecular orbital) method, *J. Comput. Chem.*, 2000, **21**(16), 1419–1432.
50 J. Zhou, P. Tao, J. F. Fisher, Q. Shi, S. Mobashery and H. B. Schlegel, QM/MM studies of the matrix metalloproteinase 2 (MMP2) inhibition mechanism of (S)-SB-3CT and its oxirane analogue, *J. Chem. Theory Comput.*, 2010, **6**(11), 3580–3587.
51 C. U. Ibeji, G. F. Tolufashe, T. Ntombela, T. Govender, G. E. Maguire and G. Lamichhane, *et al.*, The catalytic role of water in the binding site of L, D-transpeptidase 2 within acylation mechanism: a QM/MM (ONIOM) modelling, *Tuberculosis*, 2018, **113**, 222–230.

52 G. F. Tolufashe, A. K. Halder, C. U. Ibeji, M. M. Lawal, T. Ntombela and T. Govender, *et al.*, Inhibition of Mycobacterium tuberculosis L, D-Transpeptidase 5 by Carbapenems: MD and QM/MM Mechanistic Studies, *ChemistrySelect*, 2018, **3**(48), 13603–13612.

53 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo and K. M. Merz, *et al.*, The Amber biomolecular simulation programs, *J. Comput. Chem.*, 2005, **26**(16), 1668–1688.

54 H. G. Kruger, Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *THEOCHEM*, 2002, **577**(2–3), 281–285.

55 V. Gokul, H. G. Kruger, T. Govender, L. Fourie and T. D. Power, An ab initio mechanistic understanding of the regioselective acetylation of 8, 11-dihydroxy-pentacyclo [5.4. 0.02, 6.03, 10.05, 9] undecane-8, 11-lactam, *THEOCHEM*, 2004, **672**(1–3), 119–125.

56 T. Singh, H. G. Kruger, K. Bisetty and T. D. Power, Theoretical study on the formation of a pentacyclo-undecane cage lactam, *Comput. Theor. Chem.*, 2012, **986**, 63–70.

57 S. Mosebi, L. Morris, H. W. Dirr and 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, *J. Virol.*, 2008, **82**(22), 11476–11479.

58 P. Naicker, S. Stoychev, H. W. Dirr and Y. Sayed, Amide hydrogen exchange in HIV–1 subtype B and C proteases–insights into reduced drug susceptibility and dimer stability, *FEBS J.*, 2014, **281**(24), 5395–5410.

59 A. Williams, A. Basson, I. Achilonu, H. W. Dirr, L. Morris and Y. Sayed, Double trouble? Gag in conjunction with double insert in HIV protease contributes to reduced DRV susceptibility, *Biochem. J.*, 2019, **476**(2), 375–384.

60 P. Boross, P. Bagossi, T. D. Copeland, S. Oroszlan, J. M. Louis and J. Tözsér, Effect of substrate residues on the P2′ preference of retroviral proteinases, *Eur. J. Biochem.*, 1999, **264**(3), 921–929.

61 J. Tözsér, P. Bagossi, I. T. Weber, T. D. Copeland and S. Oroszlan, Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases, *J. Biol. Chem.*, 1996, **271**(12), 6781–6788.

62 A. Fehér, I. T. Weber, P. Bagossi, P. Boross, B. Mahalingam and J. M. Louis, *et al.*, Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites, *Eur. J. Biochem.*, 2002, **269**(16), 4114–4120.

63 D. R. Kipp, R. G. Silva and V. L. Schramm, Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease, *J. Am. Chem. Soc.*, 2011, **133**(48), 19358–19361.

64 B. Maschera, G. Darby, G. Palú, L. L. Wright, M. Tisdale and R. Myers, *et al.*, Human immunodeficiency virus mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex, *J. Biol. Chem.*, 1996, **271**(52), 33231–33235.

65 E. J. Rodriguez, C. Debouck, I. C. Deckman, H. Abu-Soud, F. M. Raushel and T. D. Meek, Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry, *Biochemistry*, 1993, **32**(14), 3557–3563.

66 P. Naicker, I. Achilonu, S. Fanucchi, M. Fernandes, M. A. Ibrahim and H. W. Dirr, *et al.*, Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance, *J. Biomol. Struct. Dyn.*, 2013, **31**(12), 1370–1380.

67 M. Prabu-Jeyabalan, E. Nalivaika and C. A. Schiffer, How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of HIV-1 protease1, *J. Mol. Biol.*, 2000, **301**(5), 1207–1220.

68 M. Prabu-Jeyabalan, E. Nalivaika and C. A. Schiffer, Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, *Structure*, 2002, **10**(3), 369–381.

69 K. Morokuma, New challenges in quantum chemistry: quests for accurate calculations for large molecular systems, *Philos. Trans. R. Soc., A*, 2002, **360**(1795), 1149–1164.

70 T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery and K. Morokuma, *et al.*, Combining quantum mechanics methods with molecular mechanics methods in ONIOM, *J. Chem. Theory Comput.*, 2006, **2**(3), 815–826.

71 M. Jaskolski, A. G. Tomasselli, T. K. Sawyer, D. G. Staples, R. L. Heinrikson and J. Schneider, *et al.*, Structure at 2.5-. ANG. resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor, *Biochemistry*, 1991, **30**(6), 1600–1609.

72 Z. Sanusi, T. Govender, G. Maguire, S. Maseko, J. Lin and H. Kruger, *et al.*, Investigation of the binding free energies of FDA approved drugs against subtype B and C-SA HIV PR: ONIOM approach, *J. Mol. Graphics Modell.*, 2017, **76**, 77–85.

73 A. E. Reed, L. A. Curtiss and F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, *Chem. Rev.*, 1988, **88**(6), 899–926.

74 M. Prabu-Jeyabalan, E. A. Nalivaika, N. M. King and C. A. Schiffer, Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease, *J. Virol.*, 2004, **78**(22), 12446–12454.

75 Z. Liu, Y. Wang, J. Brunzelle, I. A. Kovari and L. C. Kovari, Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease, *Protein J.*, 2011, **30**(3), 173–183.

76 M. Kumar, K. Kannan, M. Hosur, N. S. Bhavesh, A. Chatterjee and R. Mittal, *et al.*, Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations, *Biochem. Biophys. Res. Commun.*, 2002, **294**(2), 395–401.

77 W. L. DeLano, *The PyMOL molecular graphics system*, 2002.

78 J. C. Cole, C. W. Murray, J. W. M. Nissink, R. D. Taylor and R. Taylor, Comparing protein–ligand docking programs is difficult, *Proteins: Struct., Funct., Bioinf.*, 2005, **60**(3), 325–332.

79 H. Gohlke, M. Hendlich and G. Klebe, Knowledge-based scoring function to predict protein-ligand interactions, *J. Mol. Biol.*, 2000, **295**(2), 337–356.

80 M. Kontoyianni, L. M. McClellan and G. S. Sokol, Evaluation of docking performance: comparative data on docking algorithms, *J. Med. Chem.*, 2004, **47**(3), 558–565.
81 F. Sussman, M. C. Villaverde, J. L. Dominguez and U. H. Danielson, On the active site protonation state in aspartic proteases: implications for drug design, *Curr. Pharm. Des.*, 2013, **19**(23), 4257–4275.
82 A. Tripathi, M. Fornabaio, F. Spyrakis, A. Mozzarelli, P. Cozzini and G. E. Kellogg, Complexity in modeling and understanding protonation states: computational titration of HIV-1-Protease–inhibitor complexes, *Chem. Biodiversity*, 2007, **4**(11), 2564–2577.
83 A. Brik and C.-H. Wong, HIV-1 protease: mechanism and drug discovery, *Org. Biomol. Chem.*, 2003, **1**(1), 5–14.
84 S. Piana and P. Carloni, Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: an ab initio study on the free enzyme, *Proteins: Struct., Funct., Bioinf.*, 2000, **39**(1), 26–36.
85 H. Li, A. D. Robertson and J. H. Jensen, Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**(4), 704–721.
86 H. Liu, F. Müller-Plathe and W. F. van Gunsteren, A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease, *J. Mol. Biol.*, 1996, **261**(3), 454–469.
87 B. Honarparvar, T. Govender, G. E. Maguire, M. E. Soliman and H. G. Kruger, Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity, *Chem. Rev.*, 2013, **114**(1), 493–537.
88 R. Dennington, T. Keith and J. Millam, *GaussView, Version*, Semichem Inc., Shawnee Mission KS, 2009, vol. 5.
89 J. Antosiewicz, J. A. McCammon and M. K. Gilson, Prediction of pH-dependent properties of proteins, *J. Mol. Biol.*, 1994, **238**(3), 415–436.
90 A. J. Ribeiro, D. Santos-Martins, N. Russo, M. J. Ramos and P. A. Fernandes, Enzymatic flexibility and reaction rate: a QM/MM study of HIV-1 protease, *ACS Catal.*, 2015, **5**(9), 5617–5626.
91 M. Prabu-Jeyabalan, E. A. Nalivaika, K. Romano and C. A. Schiffer, Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate, *J. Virol.*, 2006, **80**(7), 3607–3616.
92 T. Masuda, Molecular dynamics simulation for the reversed power stroke motion of a myosin subfragment-1, *BioSystems*, 2015, **132**, 1–5.
93 S. Humbel, S. Sieber and K. Morokuma, The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: test for n-butane conformation and SN2 reaction: RCl+ Cl^-, *J. Chem. Phys.*, 1996, **105**(5), 1959–1967.
94 M. J. Field, P. A. Bash and M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, *J. Comput. Chem.*, 1990, **11**(6), 700–733.
95 D. Wei and D. Salahub, A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution, *Chem. Phys. Lett.*, 1994, **224**(3–4), 291–296.
96 N. Heinrich, K. B. Lipkowitz and D. B. Boyd (ed.), Reviews in computational chemistry, VCH Verlagsgesellschaft, Weinheim, New York, 1990, ISBN 3-527-27845-1, 419 Seiten, Preis: DM 176, *Berichte der Bunsengesellschaft für Physikalische Chemie*, 1992, **96**(2), 224.
97 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.*, 2004, **25**(9), 1157–1174.
98 J. W. Ochterski, Vibrational analysis in Gaussian help@ gaussian com, 1999.
99 M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, *et al.*, *Gaussian 16 Revision B. 01*, Gaussian Inc., Wallingford CT, 2016.
100 J. W. Ochterski, *Thermochemistry in Gaussian*, Gaussian Inc., 2000, pp. 1–19.
101 J. Foster and F. Weinhold, Natural hybrid orbitals, *J. Am. Chem. Soc.*, 1980, **102**(24), 7211–7218.
102 F. Weinhold and C. R. Landis, *Valency and bonding: a natural bond orbital donor–acceptor perspective*, Cambridge University Press, 2005.
103 I. Majerz, Directionality of inter-and intramolecular OHO hydrogen bonds: DFT study followed by AIM and NBO analysis, *J. Phys. Chem. A*, 2012, **116**(30), 7992–8000.
104 R. Behjatmanesh-Ardakani, NBO–NEDA, NPA, and QTAIM studies on the interactions between aza-, diaza-, and triaza-12-crown-4 (An-12-crown-4, n= 1, 2, 3) and Li +, Na +, and K+ ions, *Comput. Theor. Chem.*, 2015, **1051**, 62–71.
105 A. M. Priya, L. Senthilkumar and P. Kolandaivel, Hydrogen-bonded complexes of serotonin with methanol and ethanol: a DFT study, *Struct. Chem.*, 2014, **25**(1), 139–157.
106 Z. M. Kotena, R. Behjatmanesh-Ardakani, R. Hashim and V. M. Achari, Hydrogen bonds in galactopyranoside and glucopyranoside: a density functional theory study, *J. Mol. Model.*, 2013, **19**(2), 589–599.
107 U. Ryde, A fundamental view of enthalpy–entropy compensation, *MedChemComm*, 2014, **5**(9), 1324–1336.

Concerted hydrolysis mechanism of HIV-1 natural substrate against subtypes B and C-SA PR: Insight through Molecular Dynamics and Hybrid QM/MM studies

Zainab K. Sanusi,[a] Monsurat M. Lawal,[a] Thavendran Govender,[b] Sooraj Baijnath,[a] Tricia Naicker,[a] Glenn E. M. Maguire,[a, c], Bahareh Honarparvar,[a*] and Hendrik G. Kruger.[a*]

[a]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

[b]AnSynth PTY LTD, 498 Grove End Drive, Durban 4001, South Africa.

[c]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.

***Corresponding authors**: Prof. H. G. Kruger; kruger@ukzn.ac.za and Dr B. Honarparvar, baha.honarparvar@gmail.com; Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

Table of Contents

The HIV-1 PR cleavage sites in the Gag and Gag-Pol polyproteins.

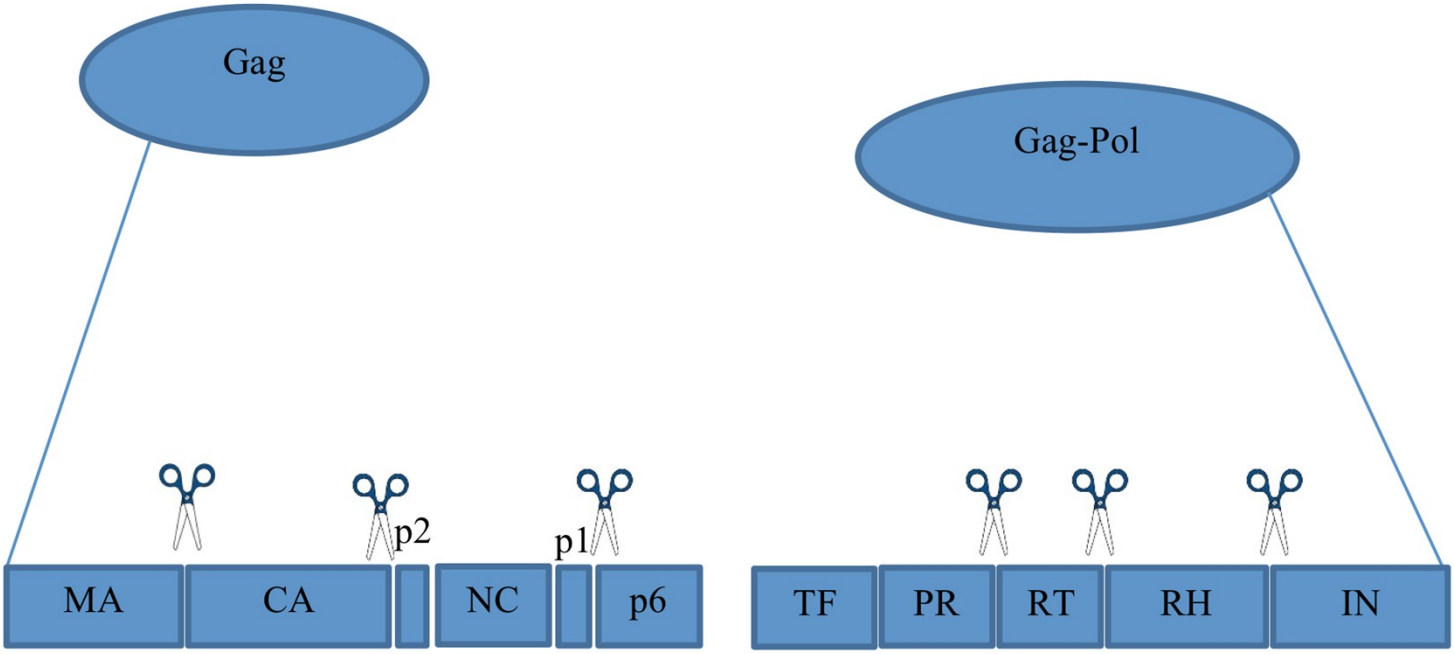


Figure S1. Schematic representation of the Gag and Gag-Pol polyproteins.

We determined (a) the standard deviation for these calculations and (b) explored whether the MD method to find starting structures for these calculations will give rise to potential conformational effects of the HIV-1 PR—substrate complexes. In order to achieve this, we performed a triplicate MD run for the PR-RT and TF-PR complexes, since these two substrates gave a reversed order for the theoretical activation energies, compared to the experimental results. We employed a random seed to create a new starting MD simulation with different atomic velocities, creating a new starting MD run with different atomic coordinates as well as different atomic velocities and changing the simulation box.[1] This was followed by a ONIOM (QM/MM) geometry optimization on the three different starting structures from the MD calculation (Table S1). The optimized ONIOM (QM/MM) TS structures for PR-RT and TF-PR enzyme—substrate complexes taken from the triplicate MD runs as well as the lowest substrate conformation taken from the restrained MD (described earlier) exhibit a similar geometry overall. Note that the standard deviation for both substrates PR-RT and TF-PR were as high as 1.0 and 0.7 kcal mol^{-1} respectively. The differences in distances are small (Figure S2) with an average RMSD of 0.82 Å, this suggests that

the reversed orders for the activation energies is not a result of different poses for the starting structures. It also gives assurance that the method to create starting structures produces consistent minimum conformations for the substrates.

Table S1. The relative thermodynamics of the three ONIOM (QM/MM) optimized MD run of PR-RT and TF-PR enzyme—substrate complexes as well as the lowest conformation using ONIOM [B3LYP/6-31++G(d,p)].

HIV-1 PR B—PR-RT	**ΔG‡**	**Average value**	**Standard deviation**
1. ONIOM after lowest conformation	14.5		
2. ONIOM after random seed in MD	14.5		
3. ONIOM after creating a different atomic coordinates and velocities in MD	15.7	15.3	1.0
4. ONIOM after changing the size of simulation box in MD	16.6		
HIV-1 PR B—TF-PR			
1. ONIOM after lowest conformation	16.5		
2. ONIOM after random seed in MD	16.6		
3. ONIOM after creating a different atomic coordinates and velocities in MD	17.5	17.1	0.7
4. ONIOM after changing the size of simulation box in MD	17.9		

Values are reported in kcal mol^{-1}, ΔG‡ =Activation free energy

The ONIOM calculation after changing the size of the simulation box[1, 2] from 10 to 12 Å should not have a drastic effect on the interaction between the natural substrate and the active sites. As mentioned earlier the average RMSD is less than 1 Å which shows good pose for the natural substrate.[3, 4]

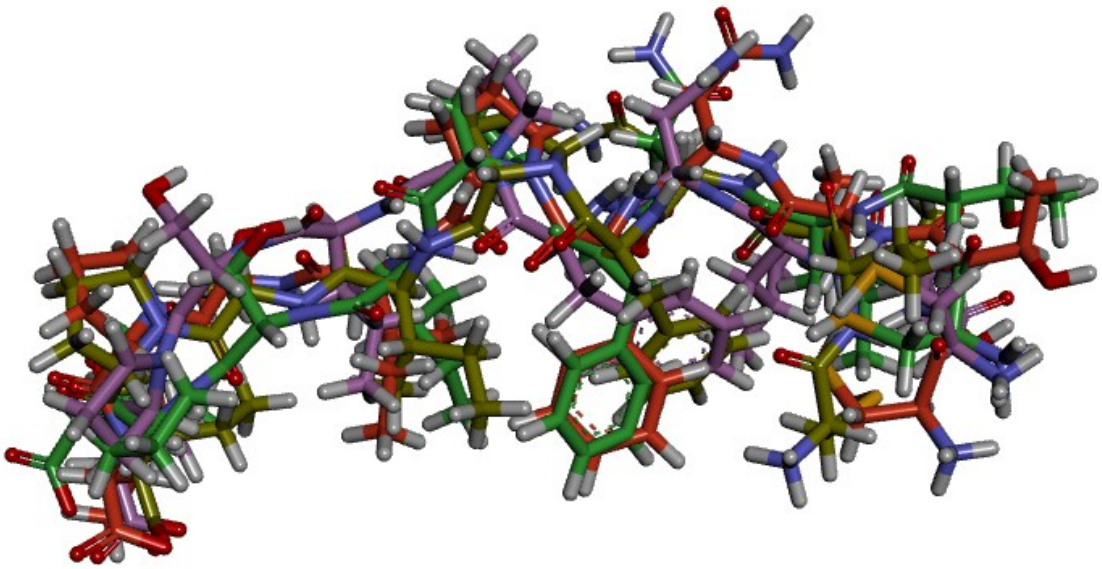


Figure S2. Superimposed PR-RT natural substrate (average RMSD of 0.82 Å) taken from triplicate MD runs showing movement and flexibility at the subtype B HIV-1 PR active site.

However, the structural characteristic of the complexes may have been altered as a result of the increase in the water residues resulting from the change in the simulation box.[5] Therefore, a virtual inspection and analysis was performed for the PR-RT and TF-PR for cases 1 and 4 (Table S1) using LigPlot[6] (see Figure S3a and S3b). As can be seen more hydrogen bond intermolecular interactions was observed for case 1 (Table S1), while case 4, after changing the size of simulation box in MD complex shows fewer hydrogen bond intermolecular interactions between the substrate and active sites. This explains why case 1 has a much lower activation energy compared to case 4 for both natural substrates Figure S3a and S3b.

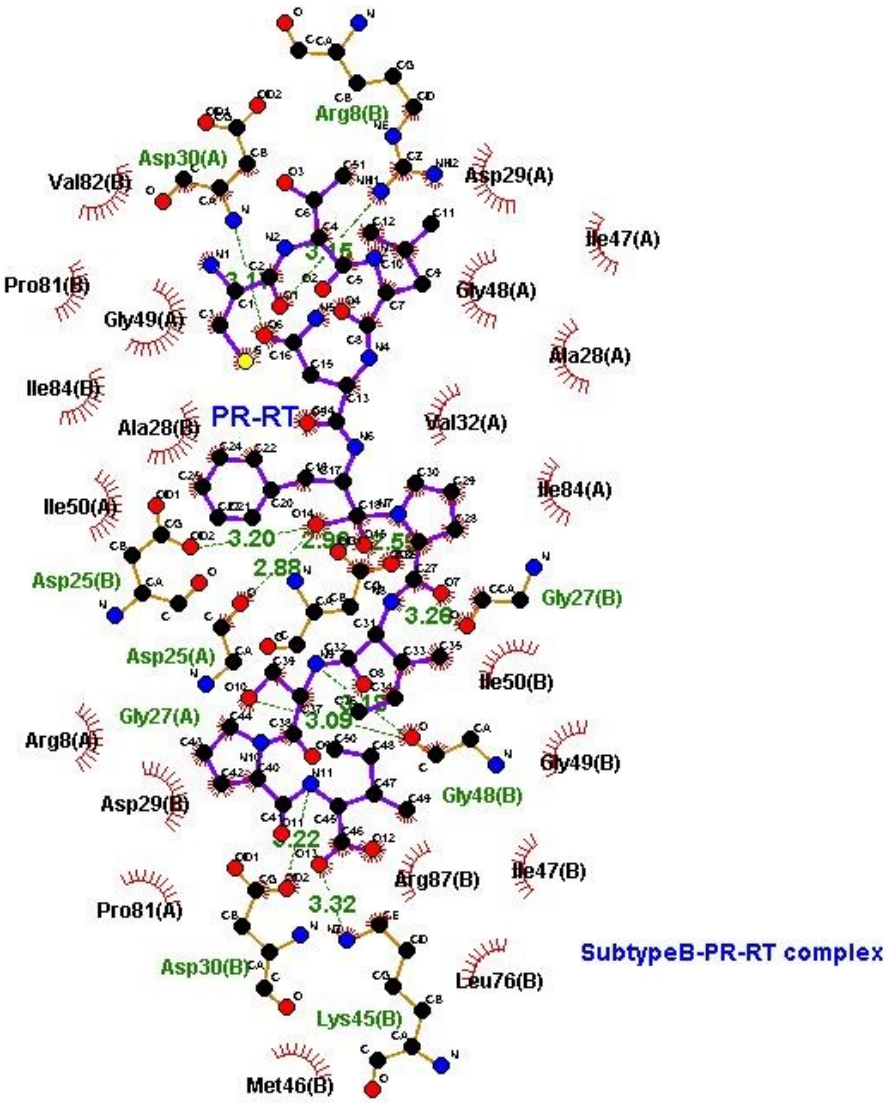


Figure S3a. Hydrogen bond interactions plot of HIV subtype B PR complexed with PR-RT natural substrate. These plots were created after ONIOM optimization of the lowest conformation (case 1) using Ligplot[6].

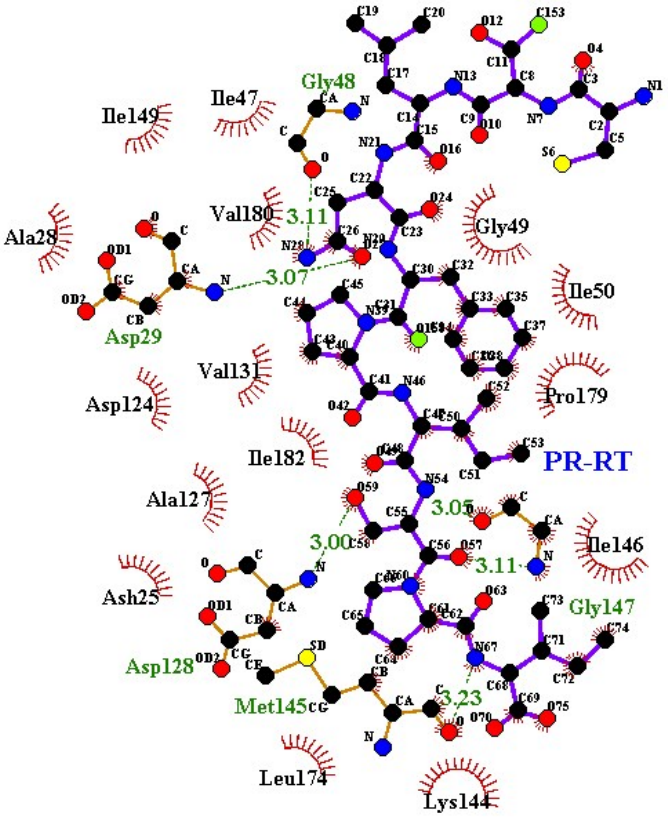


Figure S3b. Hydrogen bond interactions plot of HIV subtype B PR complexed with PR-RT natural substrate. These plots were created after ONIOM optimization of the changed simulation box to 12 Å using Ligplot[6].

Appendix: The 3D structures of all the enzyme—substrate complexes considered.

3D structures of the enzyme substrate complexes.zip

References

1. Masuda, T. (2015) Molecular dynamics simulation for the reversed power stroke motion of a myosin subfragment-1, *Biosystems 132*, 1-5.
2. Yang, L., Dordick, J. S., and Garde, S. (2004) Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity, *Biophysical Journal 87*, 812-821.
3. Cole, J. C., Murray, C. W., Nissink, J. W. M., Taylor, R. D., and Taylor, R. (2005) Comparing protein–ligand docking programs is difficult, *Proteins: Structure, Function, and Bioinformatics 60*, 325-332.

4. Kontoyianni, M., McClellan, L. M., and Sokol, G. S. (2004) Evaluation of docking performance: comparative data on docking algorithms, *Journal of Medicinal Chemistry 47*, 558-565.
5. Chong, S.-H., and Ham, S. (2013) Assessing the influence of solvation models on structural characteristics of intrinsically disordered protein, *Computational and Theoretical Chemistry 1017*, 194-199.
6. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery, *Journal of Chemical Information and Modeling 51*, 2778-2786.

Exploring the Concerted Mechanistic Pathway for HIV-1 PR—Substrate Revealed by Umbrella Sampling Simulation

Zainab K. Sanusi,[a] Monsurat M. Lawal,[a] Pancham Lal. Gupta,[d] Thavendran Govender,[b] Sooraj Baijnath,[a] Tricia Naicker,[a] Glenn E. M. Maguire,[a, c], Bahareh Honarparvar,[a] Adrian E. Roitberg,[d] and Hendrik G. Kruger.[a*]

[a]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

[b]AnSynth PTY LTD, 498 Grove End Drive, Durban 4001, South Africa.

[c]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.

[d]Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States

*__Corresponding authors__: Prof. H. G. Kruger; kruger@ukzn.ac.za

Abstract

The HIV-1 protease (HIV-1 PR) is an essential enzyme for the replication process of its virus and it is therefore considered an important target for the development of drugs against the acquired immunodeficiency syndrome (AIDs). In this paper, the concerted acyclic hydrolysis reaction catalytic mechanism by subtype B/C-SA HIV-1 PR is explored by means of enhanced sampling technique using the umbrella method. The free activation energy results were computed in terms of potential mean force (PMF) analysis within hybrid QM(DFTB)/MM approach. The theoretical findings suggest that the proposed mechanism corresponds in principle with experimental data. Given our observations, we suggest that the QM/MM MD method can be used as a reliable computational technique to rationalize lead compounds against specific targets such as the HIV-1 protease.

Keywords

HIV-1 subtypes B and C-SA PR; HIV-1 PR cleavage mechanism; Umbrella sampling method; Steered molecular dynamics (SMD); Natural substrates; Activation free energy; Concerted transition state.

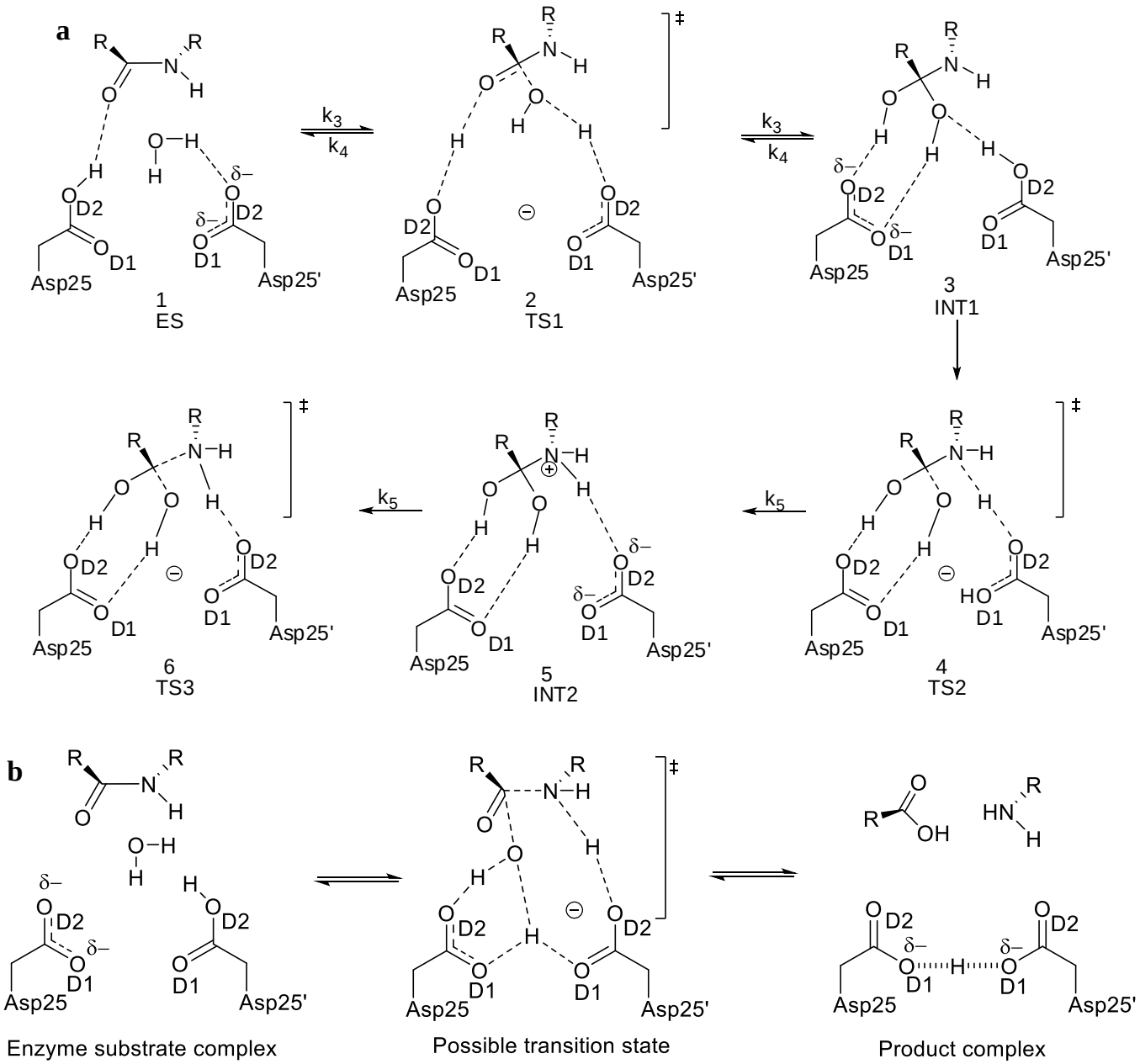
1 Introduction

In the quest for potential antiviral drugs for acquired immune deficiency syndrome (AIDs), the crystal structures of the active human immunodeficiency virus type 1 protease (HIV-1 PR) and its mutants have been investigated possibly more than any other protein structures.(Rick, Erickson, & Burt, 1998; Wlodawer & Erickson, 1993) The HIV virus encodes an aspartic protease (Asp-PR) known as the HIV PR which is essential in the replication of the infectious virus that causes AIDs.(Kohl et al., 1988; Navia et al., 1989; Veronese, DeVico, Copeland, Oroszlan, Gallo, & Sarngadharan, 1985) This enzyme catalyzes the cleavage of the viral Gag and Gag-Pol polypeptides, which are necessary for the final maturation process for the HIV life cycle.(Kohl et al., 1988; Navia et al., 1989; Veronese et al., 1985) The inhibition of this process results in the production of immature, noninfectious viral particles.(Chatfield, Eurenius, & Brooks, 1998; Roberts et al., 1990; Weber, Cavanaugh, & Harrison, 1998; Wlodawer & Erickson, 1993) Earlier experimental(Coates et al., 2006; Hyland, Tomaszek Jr, & Meek, 1991; Northrop, 2001; Rodriguez, Angeles, & Meek, 1993) and theoretical(Bjelic & Åqvist, 2006; Park, Suh, & Lee, 2000; Piana, Bucher, Carloni, & Rothlisberger, 2004; Piana, Carloni, & Parrinello, 2002; A. M. Silva, Cachau, Sham, & Erickson, 1996; Trylska, Grochowski, & McCammon, 2004) studies of the HIV-1 PR have found that transition state (TS) analogues of this enzyme-catalyzed reaction act as good inhibitors. Nevertheless, over the past decades there is yet no consensus on the mechanistic pathway for the HIV-1 PR cleavage reaction. The structure of the catalytic aspartate dyad as well as its protonation is ultimately linked to the uncertainty of the HIV-1 PR mechanism and both are still a source of many scientific discussions.(Krzemińska, Moliner, & Świderek, 2016; Lawal, Sanusi, Govender, Maguire, Honarparvar, & Kruger, 2019) However, the reaction mechanism has been classified into two major groups; the nucleophilic(Hsu, Delbaere, James, & Hofmann, 1977; Stanton, Hartsough, & Merz Jr, 1993) and the general acid-general base (GA-GB)(Antonov, Ginodman, Kapitannikov, Barshevskaya, Gurova, & Rumsh, 1978; Jaskolski et al., 1991; Lapatto et al., 1989; Park et al., 2000; Trylska et al., 2004) process, with the latter being the most upheld mechanism for aspartic proteases, especially for HIV-1 PR (Scheme 2).(Antonov et al., 1978; Fruton, 1976; Hyland et al., 1991) These proposed mechanisms have been elaborated in our reported work.(Lawal et al.,

2019; Lawal et al., 2018; Z. K. Sanusi, Lawal, Govender, Maguire, Honarparvar, & Kruger, 2019)

Generally, in the nucleophilic mechanism, the protonated catalytic aspartate act as the nucleophile(Weber et al., 1998) which attacks the carbonyl atom of the scissile peptide (C—N) bond yielding either an anhydride intermediate (stepwise nucleophile reaction, Scheme 1**a**) or a transition state that leads to the product (concerted nucleophilic reaction, Scheme 1**b**).(Boross, Bagossi, Copeland, Oroszlan, Louis, & Tözsér, 1999; Hsu et al., 1977) The general acid-base mechanism (Scheme 2) involves a catalytic water acting as the nucleophile, which attacks the carbonyl atom of the scissile peptide (C—N) at the active site of the Asp-PR.(Krzemińska et al., 2016; Park et al., 2000) The reaction mechanism proceeds with the nitrogen atom of the scissile unit that accepts a hydrogen from the protonated Asp and the catalytic water is ionized by donating its proton to the unprotonated Asp, this phase of the mechanism can either be a step-wise (Scheme 2**a**) or a concerted process (Scheme 2**b**).(Lawal et al., 2019; Trylska, Bała, Geller, & Grochowski, 2002; Trylska et al., 2004)

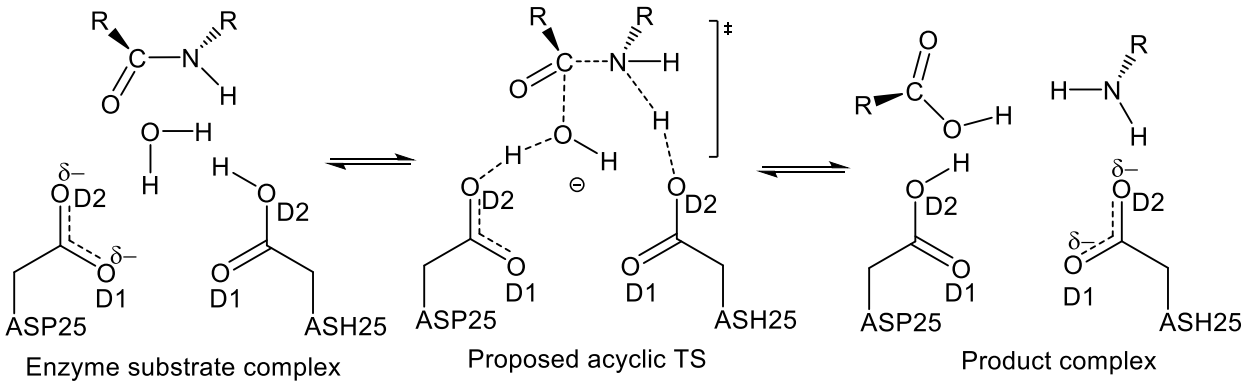
In 1991, Jaskólski *et al*(Jaskolski et al., 1991) put forward an experimentally backed perspective on the hydrolytic mechanism of HIV-1 PR and a substrate. They proposed a concerted catalytic model in which the nucleophilic water and the electrophilic aspartate (an acidic proton) attack the scissile bond in a synchronous manner. The reaction started with the acidic proton at the OD2 aspartate proximal to the N atom of the approaching scissile unit. The post-reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms.(Jaskolski et al., 1991) Based on our understanding, we have summarized this experimental process in Scheme 2**b**.



Scheme 1. Proposed (**a**) stepwise and (**b**) concerted general acid-base reaction mechanism for HIV-1 PR—substrate system, redrawn from literature.(Lawal et al., 2019) Structures along the pathway of the stepwise are indicated as follows: (1) enzyme—substrate complex; (2) water attack TS; (3) tetrahedral gem-diol intermediate; (4) scissile N-protonation TS; (5) protonated amide intermediate; and (6) cleavage of scissile C-N bond TS leading to separated product complex.

With regards to the concerted cleavage mechanism of HIV-1 PR two theoretical investigations have reported transition states with activation free energies of 30(Park et al., 2000) and 43.5(Krzemińska et al., 2016) kcal mol^{-1}. However, these are higher values in comparison to

similar experimental data found in literature. (Boross et al., 1999; Hyland et al., 1991; Kipp, Silva, & Schramm, 2011; Rodriguez, Debouck, Deckman, Abu-Soud, Raushel, & Meek, 1993; Tözsér, Bagossi, Weber, Copeland, & Oroszlan, 1996) The feasibility of the concerted general acid-base model which involves a 6-membered ring TS with either one or two water molecule(s), and an acyclic TS involving one water molecule was reported in our previous ONIOM based study.(Lawal et al., 2018) It was observed that the concerted acyclic reaction mechanism (Scheme 3) of HIV-1 PR gave an optimal basis for substrate cleavage,(Lawal et al., 2018) likewise, a longer peptide sequence was necessary for favorable recognition.(Z. K. Sanusi et al., 2019) The calculated activation free energy values from these studies(Z. K. Sanusi et al., 2020; Lawal et al., 2018; Z. K. Sanusi et al., 2019) were found to be within the experimental range.



Scheme 2. Proposed concerted acyclic general acid-base reaction mechanism for HIV-1 PR—substrate system. The main difference between Scheme 2b and our proposed TS is that the acidic Asp became basic while the de-protonated Asp becomes acidic.

Studies of the HIV-1 PR structures show conformational variances in the apo enzyme and a variety of substrate/ligand-bound enzymes,(Rick et al., 1998; Schulz, 1991) hence the energy difference between the different conformational states is an important factor. Umbrella sampling,(G. Torrie & Valleau, 1977; G. M. Torrie & Valleau, 1974) an enhanced sampling technique has been applied to study reaction rates from free-energy barriers by establishing an appropriate reaction pathway.(Rick et al., 1998) Once the path is known the potential mean force (PMF)(Hazel & Gumbart, 2017) can be determined through a series of windows and are finally combined with either umbrella integration(Kästner & Thiel, 2005) or the weighted histogram analysis method (WHAM).(S. Kumar, Rosenberg, Bouzida, Swendsen, & Kollman, 1992; Souaille & Roux, 2001)

The application of the hybrid quantum mechanics/molecular mechanics (QM/MM) combined with molecular dynamics (MD) simulations has been used to study the catalytic mechanism of enzyme—substrate and enzyme—intermediate complexes of the HIV-1 PR(Krzemińska et al., 2016) and other related biological systems.(Di Martino, Masetti, Cavalli, & Recanatini, 2014; Di Palma, Bottaro, & Bussi, 2015; J. R. A. Silva et al., 2015; J. R. r. A. Silva, Roitberg, & Alves, 2014; Tsukamoto, Sakae, Itoh, Suzuki, & Okamoto, 2018; Yildirim, Park, Disney, & Schatz, 2013) This method provides improved sampling in atomistic simulations by biasing potentials (one or more dimensional) along a reaction coordinate to drive a system from one thermodynamic state to another (reactant to transition state or product).(Kästner, 2011; Rick et al., 1998)

Similarly, the protonation state of the catalytic aspartate (25/25′) has been extensively studied,(Meek et al., 1989; Northrop, 2001; Piana & Carloni, 2000; Piana, Sebastiani, Carloni, & Parrinello, 2001; Smith, Brereton, Chai, & Kent, 1996) in fact, all ten possibilities (mono-protonation at different O2 of the Asp, di-protonation, un-protonated, low barrier hydrogen) have been considered in the past.(Lawal et al., 2019) Nevertheless, the choice of protonation herein was influenced by Meek *et al*(Hyland et al., 1991; McGee, Edwards, & Roitberg, 2014) who demonstrated that the *pKa* of one of the catalytic Asp increases to 5.2, while the other is 3.1 when bound to a substrate/inhibitor. This suggests that the catalytic aspartate group at the active site is mono-protonated.(Adachi et al., 2009; Wang et al., 1996)

The HIV-1 PR hydrolysis of the Gag and Gag-Pol polypeptide involves the cleavage and formation of chemical bonds,(Park et al., 2000) and the catalytic process in this study follows a concerted general acid-base mechanism with a mono-protonation of the catalytic aspartate at the active site.(Z. K. Sanusi et al., 2020; Z. Sanusi et al., 2017; Z. K. Sanusi et al., 2018; Z. K. Sanusi et al., 2019) Herein, a concerted acyclic TS reaction mechanism of the HIV-1 PR—substrate involving one water molecule as well as a longer natural substrate sequence (P5-P5′) from our previous ONIOM report was employed,(Z. K. Sanusi et al., 2020; Lawal et al., 2018; Z. K. Sanusi et al., 2019) the present work is aimed at providing more information of the mechanistic landscape of the HIV-1 PR system. Both the catalytic Asp (25/25′) as well as a water molecule are involved in the reaction mechanism, where the protonated Asp becomes basic by losing its proton and the unprotonated Asp gains a proton and becomes acidic.

We applied hybrid QM/MM MD simulations and umbrella sampling approaches to study the details of the concerted acyclic catalytic mechanism of the HIV-1 PR. In the QM/MM methodology, a part of the system (active site and the natural substrate) is described by quantum mechanics while the remaining part (the whole system) is described by classical mechanics. The SCC-DFTB(Elstner et al., 1998) method is used to describe the QM region of the system, this method has not only demonstrated an accurate level of theory for describing the energetics of chemical reactions,(Woodcock, Hodošček, & Brooks, 2007) but also established for large biological systems yielding minimum energy path results in good agreement with high level of theory such as MP2.(Barnett & Naidoo, 2010) To ensure that the calculation is not an artefact of a single enzyme—substrate system, two PRs, HIV-1 subtype B complexed with a natural substrate (NC-p1) is set as the standard/control and subtype C-South Africa (C-SA) HIV-1 PR complexed with three natural substrates Matrix-capsid (MA-CA), nucleopsid-p1 (NC-p1) and protease-reverse transcriptase (PR-RT) were used. Since the active enzyme as well as its mutants recognizes and cleaves the same polypeptide precursors, this study was aimed at exploring the proposed mechanism of proteolysis catalyzed by HIV-1 PR *via* QM/MM umbrella sampling technique using an adequate reaction coordinate. Figure 1 demonstrates the structure of the HIV-1 PR, showing graphical details of the catalytic Asp (25/25′) at the active site and a natural substrate. The recognition sequences cleaved by the HIV-1 PR is provided in Table S1 of the supporting information (SI).

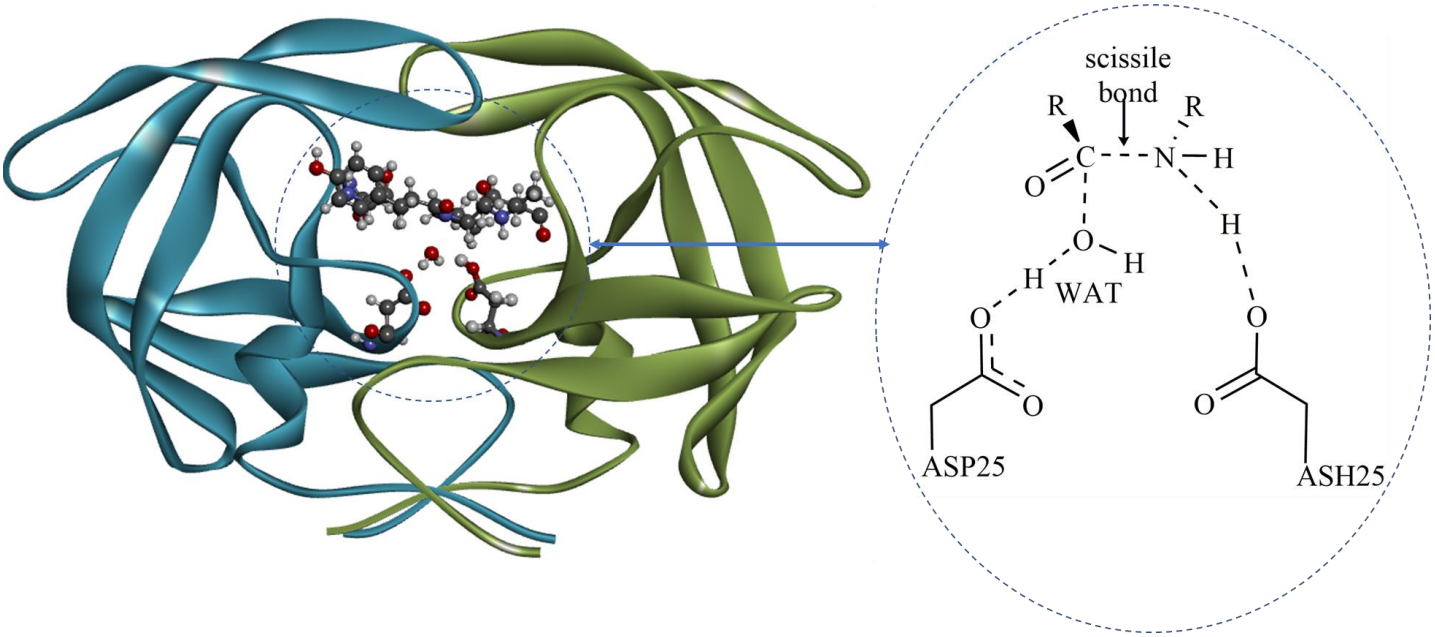


Figure 1. Graphical structure of the HIV-1 PR, showing the direction of the hydrogen at the catalytic active site, which includes the aspartates and peptide substrate. The circled-out part is the area of the system treated at QM (DFTB) level of theory, while the rest of the system is treated at the MM level of theory. All histogram and free-energy plots are provided in the supporting information (SI). The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

2 Computational methods

2.1 Structure preparation of the HIV-1 natural substrate complexes

The initial X-ray coordinates of the subtypes B and C-SA HIV-1 PR were obtained from the Protein Data Bank (PDB code: 2P3B(Kempf et al., 1990) and 3U71(Naicker et al., 2013) respectively) which contain a bound inhibitor fragment. These bound-inhibitors were exchanged for three of the HIV-1 natural substrates (MA-CA, PR-RT and NC-p1) derived also from PDB bank code: 1KJ4,(Prabu-Jeyabalan, Nalivaika, & Schiffer, 2002) 3OU3,(Z. Liu, Wang, Brunzelle, Kovari, & Kovari, 2011) 1TSU(Prabu-Jeyabalan, Nalivaika, King, & Schiffer, 2004) (Table S1) respectively. The contribution of water is an important aspect for the HIV-1 PR proteolysis, therefore, a catalytic water molecule close to the active binding site of HIV-1 PR was taken from PDB code:1LV1.(M. Kumar et al., 2002) Our model complexes used in this study were obtained by overlaying the crystal structures of both subtypes B and C-SA HIV-1 PR with the natural substrates and catalytic water using the same superimposition method explained in our previous papers.(Z. Sanusi et al., 2017; Z. K. Sanusi et al., 2018) The root mean square (RMS) values of the enzyme—substrate systems after superimposition is at the range of 1.2-1.5

Å, this was measured using PyMOL,(DeLano, 2002) which is helpful in evaluating how well the substrates are overlaid in the complex systems. An optimal superimposition is considered satisfactory if the RMS is less than 2 Å,(Cole, Murray, Nissink, Taylor, & Taylor, 2005; Gohlke, Hendlich, & Klebe, 2000; Kontoyianni, McClellan, & Sokol, 2004) hence the natural substrates considered maintain the same position at the active binding sites of the subtype B/C-SA HIV-1 PR as the substituted inhibitors. The structures of the entire enzyme—substrate systems are refined by removing any present ions and crystallographic waters from the complexes except for the catalytic water at the active site. These enzyme—substrate (ES) models serve as a starting structure for all calculations carried out in this study.

A major area of debate in modelling the HIV PR is the protonation state of the catalytic Asp (25/25′) at the active binding site and this has been discussed at length in literature.(Garrec, Sautet, & Fleurat-Lessard, 2011; Kontoyianni et al., 2004; Sussman, Villaverde, L Dominguez, & Danielson, 2013) An accurate assignment of the protonation states for the catalytic aspartates and other residues at pH 7 was performed using the empirical propKa(Li, Robertson, & Jensen, 2005) server by recalculating the standard *pKa* values of titratable amino acids. A mono-protonation model of the catalytic aspartate at the binding site was adopted and this approach is braced by the majority of related studies.(Jaskolski et al., 1991; Krzemińska et al., 2016; H. Liu, Müller-Plathe, & van Gunsteren, 1996; Park et al., 2000; Piana & Carloni, 2000) To prepare the enzyme—substrate complexes for MD simulations and the free energy calculations, the proton (H^+) of the protonated Asp in the ES system complexes is positioned midway to facilitate its donation to the nitrogen atom of the scissile peptide bond. This position has been measured to be energetically stable *via* MD simulations.(Fakhar, Govender, Lamichhane, Maguire, Kruger, & Honarparvar, 2017)

The charge of the natural substrate was determined using the restrained electrostatic potential (RESP) method by applying the HF/6-31G* level of theory using the Gaussian16 package.(Frisch et al.) Subsequently, missing atoms were added to the system complexes using the Leap module as implemented in AMBER 18 molecular dynamics package.(Case et al.) All the enzyme—substrate systems considered were solvated in an octahedral box of TIP3P(Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) water molecules, extending 12 Å outside the protein on all sides. Chlorine ions (Cl^-) were positioned around the protein to

make sure the system is neutralized, the protein and natural substrate were described using the AMBER force field 99SB.(Hornak, Abel, Okur, Strockbine, Roitberg, & Simmerling, 2006) The ES system complexes were initially energy-minimized by performing minimization at 5000 steps (of each steepest descent and conjugate gradient method) with a 20 kcal $mol^{-1}.Å^{-2}$ restraint applied to only the backbone of the protein and natural substrate peptide residues. The system is then gradually heated from 100 to 300 K using the Langevin thermostat(Pastor, Brooks, & Szabo, 1988) at a constant volume, with a restraint of 10 kcal $mol^{-1}.Å^{-2}$, over 1000 ps, and followed with 5 kcal $mol^{-1}.Å^{-2}$ restraints on the backbone atoms. This system is followed by 1000 ps of constant pressure equilibration at 300 K with 2 kcal $mol^{-1}.Å^{-2}$ restraints. All minimizations, heating and equilibration simulation steps utilized a nonbonded cutoff of 8 Å. The long-range Coulomb forces was calculated by using the particle mesh Ewald (PME) method, the time step was set to 2 fs and the SHAKE algorithm(Ryckaert, Ciccotti, & Berendsen, 1977) was used to restrain the backbone of the protein and natural substrate peptide residues during the MD simulations.

2.2 Umbrella sampling and free energy calculations

For the hybrid QM/MM MD calculations, the system was optimized where the catalytic Asp 25/25′, one water molecule at the active site and two amino acid residues of the peptide at the scissile bond were selected as the QM region that contains 62 atoms for the ES complexes, except MA-CA (63 atoms). The QM region is described using the self-consistent charge density functional tight binding (SCC-DFTB) Hamiltonian,(Seabra, Walker, Elstner, Case, & Roitberg, 2007) as implemented in AMBER18.(Case et al.) To satisfy the valence of the QM/MM boundary atoms, link atoms were placed in the substrate peptide and the catalytic aspartates' residues (Figure 1). The rest of the system (protein and water molecules) are described using the AMBER ff99SB(Hornak et al., 2006) and TIP3P(Jorgensen et al., 1983) force fields respectively, as mentioned earlier.

The free energy profile for the reaction mechanism catalyzed by HIV-1 PR (subtypes B/C-SA) was first evaluated with steered molecular dynamics (SMD) (Jarzynski, 1997a, 1997b) to drive the reaction process to completion and sample the reaction path, after which the QM/MM MD umbrella enhanced sampling was performed. The SMD method is very similar to the umbrella sampling in which the restraint center is time-dependent, and by integrating the force over

distance or time, a generalized simulation can be performed.(Case et al., 2018) In order to set up this feature for the SMD method, the NMR-specific options (nmropt=1) in the umbrella run is replaced with jar=1, which are used for generating the distance restraints based on the reference coordinates in both instances. The reaction coordinate (RC) for the acylation step that was chosen came from our previous study,(Z. K. Sanusi et al., 2020; Lawal et al., 2018; Z. K. Sanusi et al., 2019) having observed a lower energy barrier in the concerted acyclic TS reaction mechanism involving one water molecule. The RC involves the breaking of the C-N and O_{asp}-H_{asp} bonds, and the formation of HO_{wat}-C and N-Hasp bonds, hence, the acyclic TS step is defined as; RC = d(C-N) + d(O_{asp}-H_{asp}) – d(HO_{wat}-C) – d(N-H_{asp}) (see Figure 2, scheme 3).

The simulation was started by performing SMD calculation from the initial state (RC=-4.00 Å) to the final state (RC=+4.00 Å) for 50 ps. The initial configuration was generated for each window by fetching the structure every 0.75 ps. Later, we perform an umbrella sampling of 100 ps in steps of 0.12 Å on each window from the structure generated with the SMD simulation. The outputs of each MD in the sampled umbrella window serves as the initial/starting structure to perform the next one for which the value of the RC is restrained to the next position. All the QM/MM MD simulations were carried out restraining the reaction coordinate distance at a 100 kcal mol^{-1}.$Å^{-2}$ force constant of harmonic potential. The free energy curve was obtained by unbiasing the system along the reaction coordinate through variation free energy profile (vFEP) developed by Lee *et al*.(Lee, Radak, Huang, Wong, & York, 2013; Lee, Radak, Pabis, & York, 2012)

3 Results and discussion

3.1 The concerted acyclic general acid-base catalytic mechanism of HIV-1 PR

Recent computational research(Z. K. Sanusi et al., 2020; Lawal et al., 2019; Lawal et al., 2018; Z. K. Sanusi et al., 2019) from our group has buttressed the feasibility of the earlier suggested(Jaskolski et al., 1991) existence of the concerted process (Scheme 2**b**). Based on our previous theoretical investigations,(Z. K. Sanusi et al., 2020; Lawal et al., 2018; Z. K. Sanusi et al., 2019) we have applied umbrella sampling algorithm to explore the proposed concerted process (Scheme 3). The outcome of which follows the general acid-base model in a synchronous mechanism in which water donates one of its protons to the unprotonated Asp25

and the protonated Asp25 loses its proton to the scissile nitrogen atom; thus, the initial acidic Asp becomes basic and vice versa. Meanwhile, the nucleophilic water (OH) attacks the scissile carbon resulting in substrate cleavage (Scheme 3). In the concerted acylation mechanism (Scheme 3, Figure 2), generally, three major types of conformational state occurred. One is the formation of the HO_{wat}-C bond, the $d(HO_{wat}\text{-}C)$ decreases considerably from 3.40 Å in ES to 2.61 Å in TS, and then to 1.30 Å in product complex (PC). This indicates the possibility of the nucleophilic attack of the OH group of water to the carbonyl carbon (C) of the peptide at the scissile bond.

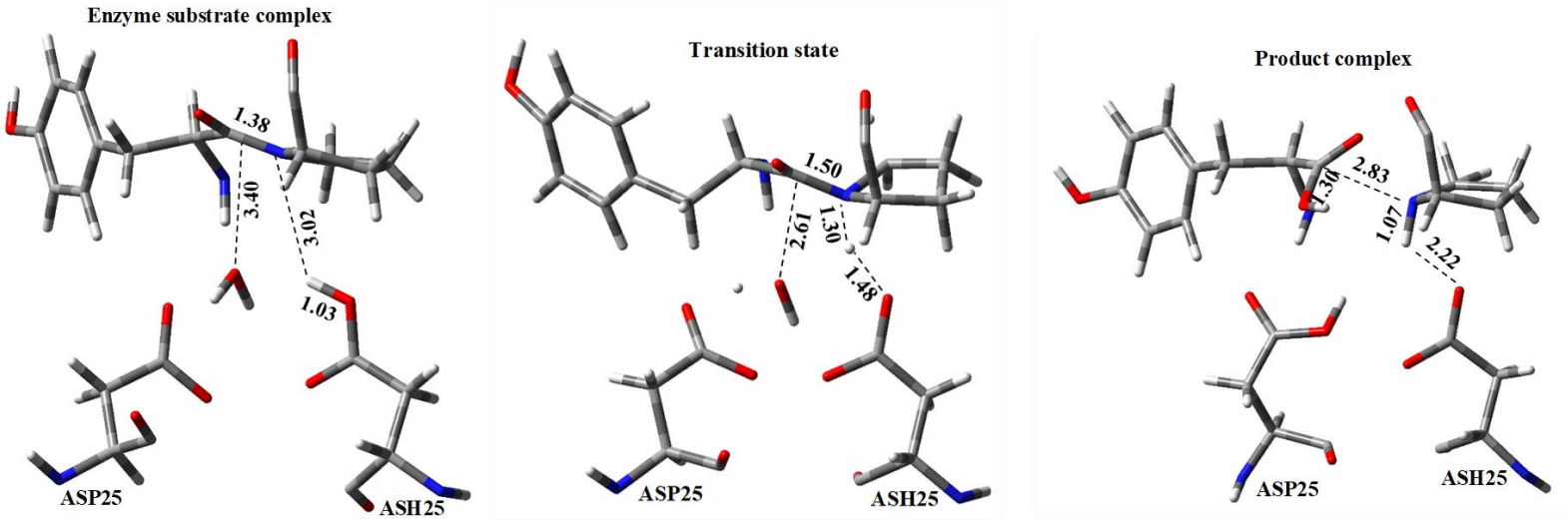


Figure 2. Proposed concerted acyclic general acid-base reaction mechanism for HIV-1 PR—substrate system. Showing the distances in the chosen reaction coordinates, from the ES to TS and product complex for one of the enzymes—substrate complex. Values are reported in Å.

Another important conformational change is the cleavage of the C-N bond, the d(C-N) increases from 1.38 Å in ES to 1.50 Å in transition state model, and then to 2.83 Å in PC, which indicates the breaking of the natural substrate's peptide bond. The last event is the proton transfer from the protonated aspartate (H_{asp}) to the nitrogen atom (N) of the substrate at the scissile bond, the $d(\text{N-}H_{asp})$ and $d(O_{asp}\text{-}H_{asp})$ respectively changes from 3.02 and 1.03 Å in ES to 1.30 and 1.48 Å in TS, and then to 1.07 and 2.22 Å in PC. The measured distances are summarized in Table **1,** S**2**, S**3**, and S**4** for all the enzyme—substrate complexes considered.

Table **1**. Averaged distances for ES, TS and PC structures for the concerted mechanism obtained by DFTB/MM 1D-PMF (Umbrella sampling) for subtype C-SA—MA-CA complex.

DFTB/MM	**ES**	**TS**	**PC**
$d(HO_{wat}\text{-}C)$	3.40	2.61	1.30
$d(C\text{-}N)$	1.38	1.50	2.83
$d(N\text{-}H_{asp})$	3.02	1.30	1.07
$d(O_{asp}\text{-}H_{asp})$	1.03	1.48	2.22

Values are reported in Å. Provided in the supplementary information is the table for the average distances for ES, TS and PC of the remaining considered complexes. All histogram and free-energy plots are provided in the supporting information (SI). The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

The plausibility of this mechanistic process was evaluated by calculating the activation free energy ($\Delta G^{\ddagger}$) value associated with each transition state (TS) structure (Table **2**) of the respective complex systems. It was noticed that the $\Delta G^{\ddagger}$ obtained from the umbrella sampling approach are between 14.2 and 24.7 kcal mol^{-1}. The calculated values follow the same trend as experimental results(Fehér et al., 2002; Lawal et al., 2019; Maschera et al., 1996) and previously calculated values(Z. K. Sanusi et al., 2020) in which an ONIOM approach was used. The overall reaction process (Figures 3**a** and 4**a**) is exergonic with negative free energy values for the product complexes (PC).

Table **1**. Free energy value of the concerted acyclic general acid-base cleavage mechanism of natural substrate by HIV-1 PR (subtypes B and C-SA) using QM(DFTB)/MM MD (Umbrella sampling)

	B—NC-p1	**C-SA—MA-CA**	**C-SA—NC-p1**	**C-SA—PR-RT**
$\Delta G^{\ddagger}$	18.3	14.2	24.7	15.5
PC	-8.0	-18.0	-4.6	-2.5
$\Delta G^{o\ddagger}$	19.7	13.5	22.6	16.4
$\Delta G^{e\ddagger}$	18.7	15.7	18.7	15.8

Values are reported in kilocalories per mole (kcal mol^{-1}).
$\Delta G^{\ddagger}$ = calculated activation free energy (Umbrella sampling), PC product complex, $\Delta G^{o\ddagger}$ = calculated activation free energy from previous paper using ONIOM approach,(Z. K. Sanusi et al., 2020) $\Delta Ge^{\ddagger}$ = experimental activation free energy from literature for subtype B HIV-1 PR.(Fehér et al., 2002; Lawal et al., 2019; Maschera et al., 1996) All histogram and free-energy plots are provided in the supporting information (SI). The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

3.2 Free energy analysis

It should be noted that since the active enzyme cleaves the same scissile precursor as its mutant, the computed results (Table **2**) were compared with respect to the activation free energies of the HIV-1 PR found in literature.(Fehér et al., 2002; Lawal et al., 2019; Maschera et al., 1996) For the subtype B—NC-p1 system, $\Delta G^{\ddagger}$ of 18.3 kcal mol^{-1} was obtained for the cleavage of the scissile bond Phe*Asn (Table **2**, Figure 3**b**), which is -0.4 kcal mol^{-1} lower than experimental HIV-1 PR hydrolysis data of the same substrate.(Fehér et al., 2002; Lawal et al., 2019; Maschera et al., 1996)

However, the subtype C-SA—NC-p1 system gave a higher value of 24.7 kcal mol^{-1}, approximately 6 and 2.1 kcal mol^{-1} greater than experimental (subtype B HIV-1 PR) and previously calculated data(Z. K. Sanusi et al., 2020) respectively. This could be as a result of the eight-point mutation that occurs in the subtype C-SA HIV-1 PR. Likewise, NC-p1 natural substrate has been suggested to have the slowest mechanistic process as well as being the rate-limiting cleavage step in the maturation of Gag.(Prabu-Jeyabalan, Nalivaika, Romano, & Schiffer, 2006) It is noticeable from all components in Table **2** that this system (C-SA—NC-p1) produced the highest $\Delta G^{\ddagger}$ values for both calculated theoretical methods. Hence, it could be that our calculated result is accurate but needs experimental verification for HIV-1 subtype C-SA PR. From the analysis of the calculated $\Delta G^{\ddagger}$ obtained through enhanced sampling approach herein, it is tenable to say that the calculated values are comparable with experimental results and the theoretical method describes the proposed concerted mechanism reasonably.

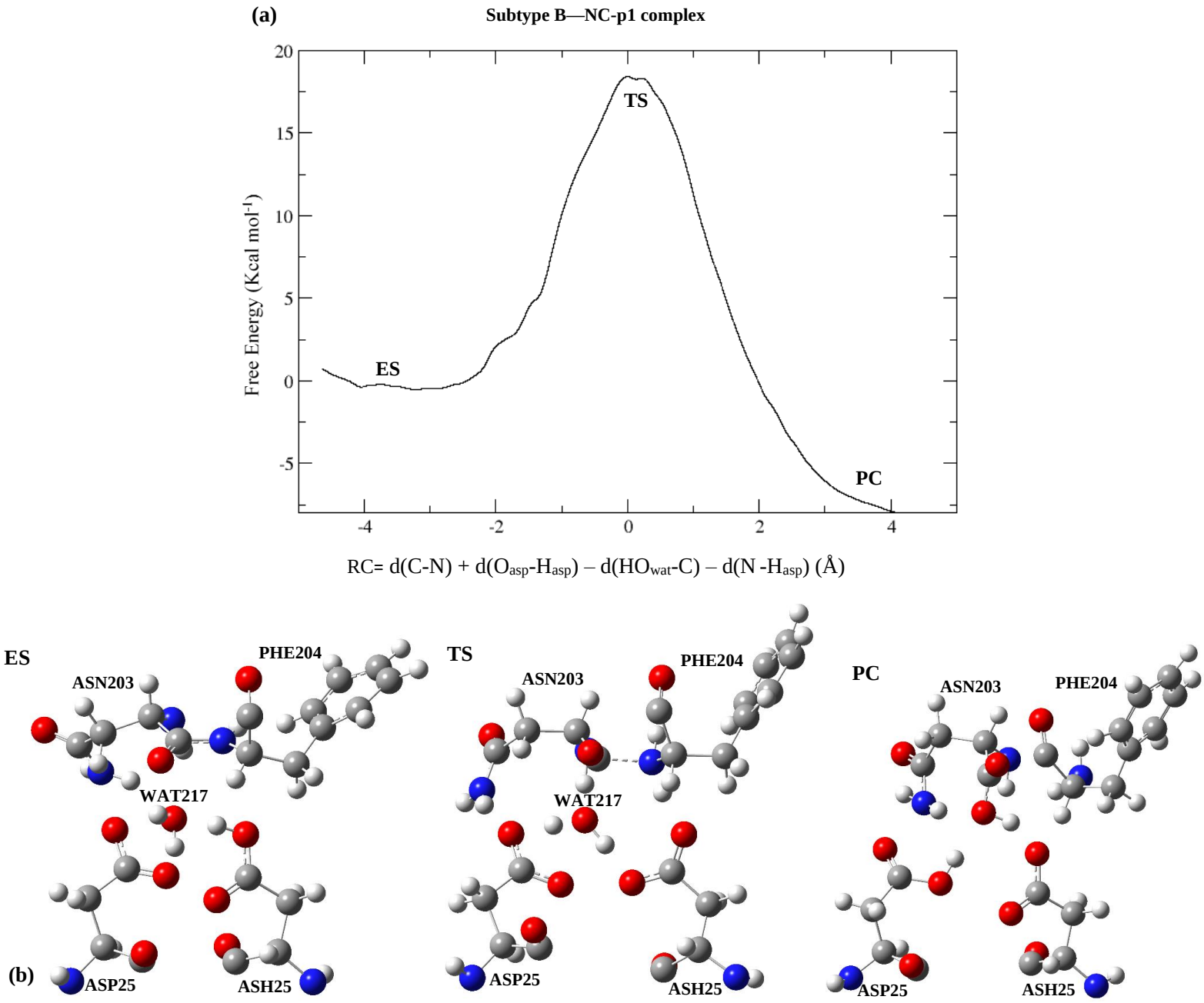


Figure 3. (**a**) Free energy profile for the concerted acyclic mechanism of subtype B HIV-1 PR with NC-p1 natural substrate. (**b**) 3D structures of enzyme—substrate (ES), transition state (TS), and product complex (PC) for the acyclic mechanistic process taken from trajectories. All histogram and free-energy plots are provided in the supporting information (SI). The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

The product complexes gave free energy values within -18 to -2.5 kcal mol^{-1} (Table **2**), thus indicating that the entire mechanism is energetically favorable. Provided in Figure 3**a** is the free energy pathway involving subtype B catalysis of NC-p1 indicating the scanned reaction

coordinates for the acyclic TS model. The zoomed in structural details of the reaction center is provided in Figure 3**b**. The free energy profile for the concerted acyclic mechanism of subtype C-SA HIV-1 PR cleavage of MA-CA natural substrate is also provided in Figure 4**a** with the 3D moieties along its reaction path in Figure 4**b**.

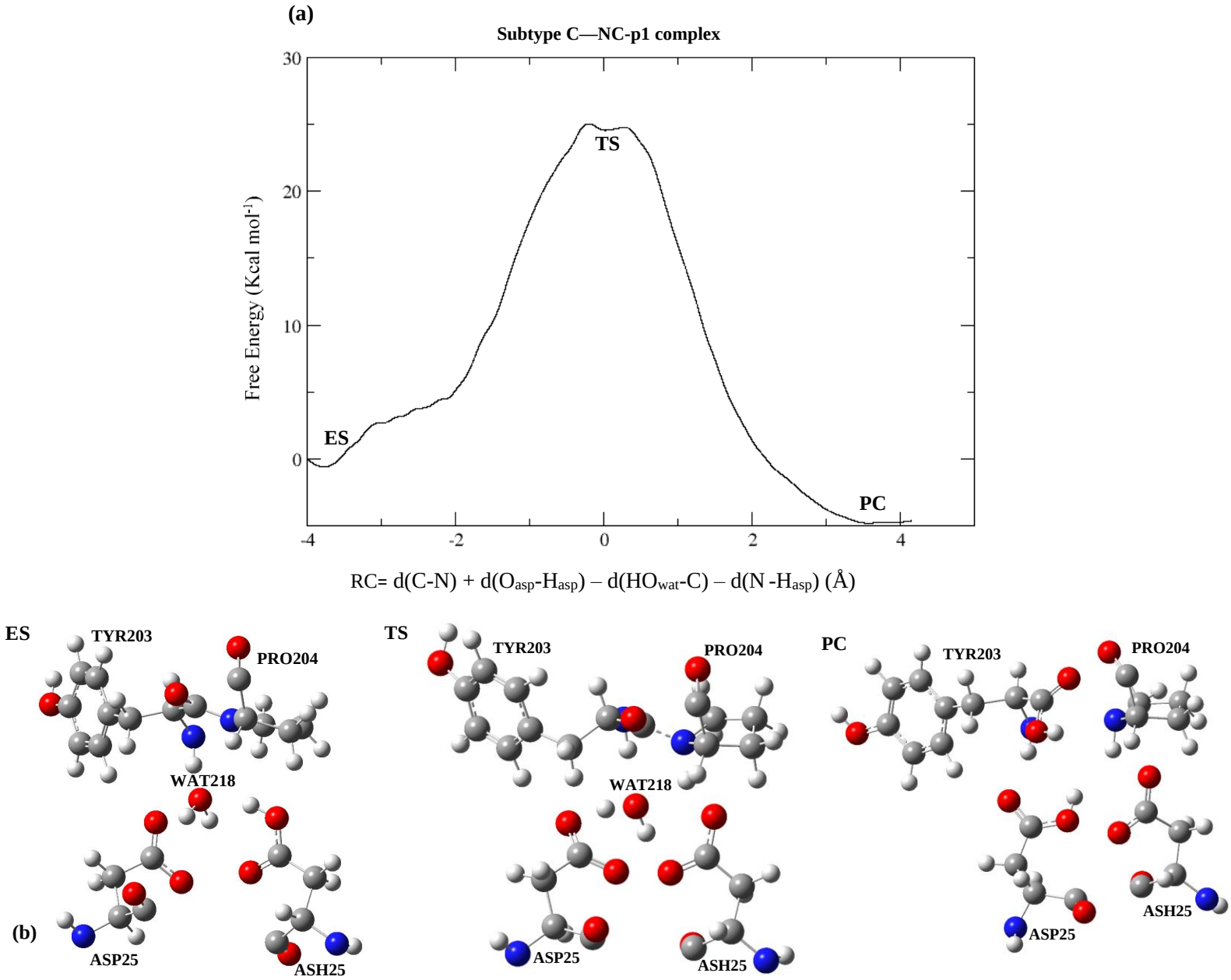


Figure 4. (**a**) Free energy profile for the concerted acyclic mechanism of subtype C-SA HIV-1 PR with MA-CA natural substrate. (**b**) 3D enzyme—substrate (ES), transition state (TS), and product complex (PC) for the acyclic mechanistic process taken from trajectories. All histogram and free-energy plots are provided in the supporting information (SI). The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

4 Conclusion

The enzymatic mechanism catalyzed by HIV-1 PR in the presence of its natural substrate is the focus of this study, and this is described by means of enhanced umbrella sampling calculations using a hybrid DFTB/MM level of theory. We envisaged that the QM/MM MD method would provide detailed information on the proposed concerted mechanism and based on the results obtained we were able to compute a complete profile of the possible reaction pathway that corresponds to the concerted acylation mechanism. The reaction process starts with a nucleophilic attack from the catalytic water on the carbonyl atom of the scissile bond, which is synchronously followed by a proton transfer from the protonated catalytic Asp25 to the nitrogen atom of the scissile bond. The cleavage of the peptide bond as a result of the nucleophilic attack from the water and hydrogen transfer from the water molecule to the unprotonated aspartate also occurs concurrently. The present investigation provides a single rate-limiting step with computed free energy barrier of 14.2, 24.7 (18.3 for subtype B—NC-p1) and 15.5 kcal mol^{-1} for the cleavage of MA-CA, NC-p1 and PR-RT by subtype C-SA, respectively. These calculated energies correspond well with the experimental data for the above natural substrates catalyzed by HIV-1 protease. Since these calculations require moderately high computational cost (long MD simulations), we were restricted to performing a 1-dimensional calculation, however, the accuracy of the results (in comparison with experimental results) is reasonable. From our work the QM/MM MD umbrella sampling is a viable method for calculating accurate results. The approach requires fewer parameters, convergence is easy to achieve with minimal or no error and the reaction pathway can be easily determined. However, the system set-up, appropriate choice of the RC along which the simulation is biased and can be difficult to ascertain. While the ONIOM calculation in principle has a straightforward calculation setup, the number of input parameters and errors makes achieving convergence more difficult and time consuming. Nevertheless, both theoretical techniques have proved to give results that are comparable to experimental data.

Most HIV-1 PR inhibitors are developed to resemble the molecular structure that occur during the chemical process, particularly the transition state model. We believe electronic modification of the amide bond is possible (for example via N-methylation)(Rauf et al., 2015) and the QM/MM MD umbrella sampling technique will be useful to determine if the activation energy

for hydrolysis increases. If the activation energy of the "scissile" bond of a natural peptide can be sufficiently increased, it may potentially be used as substrate-based inhibitors (with limited toxicity and side effects) for HIV-1 PR and other enzymes.

Competing interests

The authors declare that they have no competing interests.

Supporting Information

Table S1, histogram and free-energy plots are provided in the supporting information. The ES, TSs and PC structures are also provided in PDB format and these are listed in the table of content in the SI.

Acknowledgements

The authors thank College of Health Sciences, University of KwaZulu-Natal, Asphen Pharmacare, Medical Research Council and the National Research Foundation (all in South Africa) for financial support. The authors also acknowledge the University of Florida Research Computing for providing computational resources and support that contributed to the research results reported in this publication. GK thank AR for hosting ZS and ML at UF during the project. We are also grateful to the Centre for High Performance Computing (www.chpc.ac.za) for computational resources.

References

Adachi, M., Ohhara, T., Kurihara, K., Tamada, T., Honjo, E., Okazaki, N., . . . Matsumura, H. (2009). Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. *Proceedings of the National Academy of Sciences, 106*(12), 4641-4646.

Antonov, V., Ginodman, L., Kapitannikov, Y. V., Barshevskaya, T., Gurova, A., & Rumsh, L. (1978). Mechanism of pepsin catalysis: General base catalysis by the active-site carboxylate ion. *FEBS letters, 88*(1), 87-90.

Barnett, C. B., & Naidoo, K. J. (2010). Ring puckering: a metric for evaluating the accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM simulations. *The Journal of Physical Chemistry B, 114*(51), 17142-17154.

Bjelic, S., & Åqvist, J. (2006). Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry, 45*(25), 7709-7723.

Boross, P., Bagossi, P., Copeland, T. D., Oroszlan, S., Louis, J. M., & Tözsér, J. (1999). Effect of substrate residues on the P2′ preference of retroviral proteinases. *European journal of biochemistry, 264*(3), 921-929.

Case, D., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham III, T., Cruzeiro, V., . . . Gilson, M. AMBER 2018; 2018. *University of California, San Francisco*.

Case, D., Cerutti, D., Cheatham III, T., Darden, T., Duke, R., Giese, T., . . . Homeyer, N. (2018). AMBER reference manual. *University of California*.

Chatfield, D. C., Eurenius, K. P., & Brooks, B. R. (1998). HIV-1 protease cleavage mechanism: A theoretical investigation based on classical MD simulation and reaction path calculations using a hybrid QM/MM potential. *Journal of Molecular Structure: THEOCHEM, 423*(1), 79-92.

Coates, L., Erskine, P. T., Mall, S., Gill, R., Wood, S. P., Myles, D. A., & Cooper, J. B. (2006). X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *European Biophysics Journal, 35*(7), 559-566.

Cole, J. C., Murray, C. W., Nissink, J. W. M., Taylor, R. D., & Taylor, R. (2005). Comparing protein–ligand docking programs is difficult. *Proteins: Structure, Function, and Bioinformatics, 60*(3), 325-332.

DeLano, W. L. (2002). PyMol: An open-source molecular graphics tool. CCP4 Newsletter on Protein Crystallography, *40*, 82-92.

Di Martino, G. P., Masetti, M., Cavalli, A., & Recanatini, M. (2014). Mechanistic insights into P in1 peptidyl-prolyl cis-trans isomerization from umbrella sampling simulations. *Proteins: Structure, Function, and Bioinformatics, 82*(11), 2943-2956.

Di Palma, F., Bottaro, S., & Bussi, G. (2015). Kissing loop interaction in adenine riboswitch: insights from umbrella sampling simulations. *BMC bioinformatics, 16*(9), S6.

Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., . . . Seifert, G. (1998). Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B, 58*(11), 7260.

Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., Kruger, H. G., & Honarparvar, B. (2017). Computational model for the acylation step of the β-lactam ring: Potential application for l, d-transpeptidase 2 in mycobacterium tuberculosis. *Journal of Molecular Structure, 1128*, 94-102.

Fehér, A., Weber, I. T., Bagossi, P., Boross, P., Mahalingam, B., Louis, J. M., . . . Tözsér, J. (2002). Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *European journal of biochemistry, 269*(16), 4114-4120.

Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., . . . Nakatsuji, H. Gaussian 16 Revision B. 01. 2016; Gaussian Inc. *Wallingford CT*.

Fruton, J. S. (1976). The mechanism of the catalytic action of pepsin and related acid proteinases. *Advances in Enzymology and Related Areas of Molecular Biology, Volume 44*, 1-36.

Garrec, J., Sautet, P., & Fleurat-Lessard, P. (2011). Understanding the HIV-1 protease reactivity with DFT: what do we gain from recent functionals? *The Journal of Physical Chemistry B, 115*(26), 8545-8558.

Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology, 295*(2), 337-356.

Hazel, A., & Gumbart, J. C. (2017). Methods for calculating Potentials of Mean Force.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics, 65*(3), 712-725.

Hsu, I.-N., Delbaere, L. T., James, M. N., & Hofmann, T. (1977). Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 Å and sequence homology with porcine pepsin. *Nature, 266*(5598), 140.

Hyland, L. J., Tomaszek Jr, T. A., & Meek, T. D. (1991). Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry, 30*(34), 8454-8463.

Jarzynski, C. (1997a). Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E, 56*(5), 5018.

Jarzynski, C. (1997b). Nonequilibrium equality for free energy differences. *Physical Review Letters, 78*(14), 2690.

Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., . . . Wlodawer, A. (1991). Structure at 2.5-. ANG. resolution of chemically synthesized Human Immunodeficiency Virus Type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry, 30*(6), 1600-1609.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics, 79*(2), 926-935.

Kästner, J. (2011). Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 1*(6), 932-942.

Kästner, J., & Thiel, W. (2005). Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method:"Umbrella integration". *The Journal of chemical physics, 123*(14), 144104.

Kempf, D. J., Norbeck, D. W., Codacovi, L., Wang, X. C., Kohlbrenner, W. E., Wideburg, N. E., . . . Vasavanonda, S. (1990). Structure-based, C2 symmetric inhibitors of HIV protease. *Journal of medicinal chemistry, 33*(10), 2687-2689.

Kipp, D. R., Silva, R. G., & Schramm, V. L. (2011). Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease. *Journal of the American Chemical Society, 133*(48), 19358-19361.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R., . . . Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of the National Academy of Sciences, 85*(13), 4686-4690.

Kontoyianni, M., McClellan, L. M., & Sokol, G. S. (2004). Evaluation of docking performance: comparative data on docking algorithms. *Journal of medicinal chemistry, 47*(3), 558-565.

Krzemińska, A., Moliner, V., & Świderek, K. (2016). Dynamic and electrostatic effects on the reaction catalyzed by HIV-1 protease. *Journal of the American Chemical Society, 138*(50), 16283-16298.

Kumar, M., Kannan, K., Hosur, M., Bhavesh, N. S., Chatterjee, A., Mittal, R., & Hosur, R. (2002). Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. *Biochemical and biophysical research communications, 294*(2), 395-401.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry, 13*(8), 1011-1021.

Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., . . . Geoghegan, K. F. (1989). X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature, 342*(6247), 299.

Lawal, M. M., Sanusi, Z. K., Govender, T., Maguire, G. E., Honarparvar, B., & Kruger, H. G. (2019). From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets. *Current medicinal chemistry, 26, 1-34*.

Lawal, M. M., Sanusi, Z. K., Govender, T., Tolufashe, G. F., Maguire, G. E., Honarparvar, B., & Kruger, H. G. (2018). Unraveling the concerted catalytic mechanism of the human immunodeficiency virus type 1 (HIV-1) protease: a hybrid QM/MM study. *Structural Chemistry, 30*, 409-417.

Lee, T.-S., Radak, B. K., Huang, M., Wong, K.-Y., & York, D. M. (2013). Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *Journal of chemical theory and computation, 10*(1), 24-34.

Lee, T.-S., Radak, B. K., Pabis, A., & York, D. M. (2012). A new maximum likelihood approach for free energy profile construction from molecular simulations. *Journal of chemical theory and computation, 9*(1), 153-164.

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics, 61*(4), 704-721.

Liu, H., Müller-Plathe, F., & van Gunsteren, W. F. (1996). A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. *Journal of molecular biology, 261*(3), 454-469.

Liu, Z., Wang, Y., Brunzelle, J., Kovari, I. A., & Kovari, L. C. (2011). Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease. *The protein journal, 30*(3), 173-183.

Maschera, B., Darby, G., Palú, G., Wright, L. L., Tisdale, M., Myers, R., . . . Furfine, E. S. (1996). Human immunodeficiency virus mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. *Journal of Biological Chemistry, 271*(52), 33231-33235.

McGee, T. D., Jr., Edwards, J., & Roitberg, A. E. (2014). pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State. *Journal of Physical Chemistry B, 118*(44), 12577-12585.

Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., . . . Debouck, C. (1989). Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. *Proceedings of the National Academy of Sciences, 86*(6), 1841-1845.

Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. A., Dirr, H. W., . . . Sayed, Y. (2013). Structural insights into the South African HIV-1 subtype C protease: impact of hinge region dynamics and flap flexibility in drug resistance. *Journal of Biomolecular Structure and Dynamics, 31*(12), 1370-1380.

Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., . . . Springer, J. P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature, 337*(6208), 615.

Northrop, D. B. (2001). Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Accounts of chemical research, 34*(10), 790-797.

Park, H., Suh, J., & Lee, S. (2000). Ab initio studies on the catalytic mechanism of aspartic proteinases: nucleophilic versus general acid/general base mechanism. *Journal of the American Chemical Society, 122*(16), 3901-3908.

Pastor, R. W., Brooks, B. R., & Szabo, A. (1988). An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics, 65*(6), 1409-1419.

Piana, S., Bucher, D., Carloni, P., & Rothlisberger, U. (2004). Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations. *The Journal of Physical Chemistry B, 108*(30), 11139-11149.

Piana, S., & Carloni, P. (2000). Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins: Structure, Function, and Bioinformatics, 39*(1), 26-36.

Piana, S., Carloni, P., & Parrinello, M. (2002). Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *Journal of molecular biology, 319*(2), 567-583.

Piana, S., Sebastiani, D., Carloni, P., & Parrinello, M. (2001). Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *Journal of the American Chemical Society, 123*(36), 8730-8737.

Prabu-Jeyabalan, M., Nalivaika, E., & Schiffer, C. A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure, 10*(3), 369-381.

Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., & Schiffer, C. A. (2004). Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *Journal of virology, 78*(22), 12446-12454.

Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., & Schiffer, C. A. (2006). Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. *Journal of virology, 80*(7), 3607-3616.

Rauf, S. M. A., Arvidsson, P. I., Albericio, F., Govender, T., Maguire, G. E., Kruger, H. G., & Honarparvar, B. (2015). The effect of N-methylation of amino acids (Ac-X-OMe) on solubility and conformation: a DFT study. *Organic & biomolecular chemistry, 13*(39), 9993-10006.

Rick, S. W., Erickson, J. W., & Burt, S. K. (1998). Reaction path and free energy calculations of the transition between alternate conformations of HIV-1 protease. *Proteins: Structure, Function, and Bioinformatics, 32*(1), 7-16.

Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., . . . Krohn, A. (1990). Rational design of peptide-based HIV proteinase inhibitors. *Science, 248*(4953), 358-361.

Rodriguez, E. J., Angeles, T. S., & Meek, T. D. (1993). Use of nitrogen-15 kinetic isotope effects to elucidate details of the chemical mechanism of human immunodeficiency virus 1 protease. *Biochemistry, 32*(46), 12380-12385.

Rodriguez, E. J., Debouck, C., Deckman, I. C., Abu-Soud, H., Raushel, F. M., & Meek, T. D. (1993). Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry. *Biochemistry, 32*(14), 3557-3563.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics, 23*(3), 327-341.

Sanusi, Z., Govender, T., Maguire, G., Maseko, S., Lin, J., Kruger, H., & Honarparvar, B. (2017). Investigation of the binding free energies of FDA approved drugs against subtype B and C-SA HIV PR: ONIOM approach. *Journal of Molecular Graphics and Modelling, 76*, 77-85.

Sanusi, Z. K., Govender, T., Maguire, G. E., Maseko, S. B., Lin, J., Kruger, H. G., & Honarparvar, B. (2018). An insight to the molecular interactions of the FDA approved HIV PR drugs against L38L↑ N↑ L PR mutant. *Journal of computer-aided molecular design, 32*(3), 459-471.

Sanusi, Z. K., Lawal, M. M., Govender, T., Maguire, G. E., Honarparvar, B., & Kruger, H. G. (2019). Theoretical Model for HIV-1 PR That Accounts for Substrate Recognition and Preferential Cleavage of Natural Substrates. *The Journal of Physical Chemistry B, 123*(30), 6389-6400.

Sanusi, Z. K., Lawal, M. M., Govender, T., Baijnath, S., Naicker, T., Maguire, G. E., . . . Kruger, H. G. (2020). Concerted hydrolysis mechanism of HIV-1 natural substrate against subtypes B and C-SA PR: Insight through Molecular Dynamics and Hybrid QM/MM studies. *Physical Chemistry Chemical Physics, 22*, 2530-2539.

Schulz, G. E. (1991). Domain motions in proteins. *Current Opinion in Structural Biology, 1*(6), 883-888.

Seabra, G. d. M., Walker, R. C., Elstner, M., Case, D. A., & Roitberg, A. E. (2007). Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. *The Journal of Physical Chemistry A, 111*(26), 5655-5664.

Silva, A. M., Cachau, R. E., Sham, H. L., & Erickson, J. W. (1996). Inhibition and catalytic mechanism of HIV-1 aspartic protease. *Journal of molecular biology, 255*(2), 321-340.

Silva, J. R. A., Govender, T., Maguire, G. E., Kruger, H. G., Lameira, J., Roitberg, A. E., & Alves, C. N. (2015). Simulating the inhibition reaction of Mycobacterium tuberculosis L, D-transpeptidase 2 by carbapenems. *Chemical Communications, 51*(63), 12560-12562.

Silva, J. R. r. A., Roitberg, A. E., & Alves, C. u. N. (2014). Catalytic mechanism of L, D-transpeptidase 2 from Mycobacterium tuberculosis described by a computational approach: insights for the design of new antibiotics drugs. *Journal of chemical information and modeling, 54*(9), 2402-2410.

Smith, R., Brereton, I. M., Chai, R. Y., & Kent, S. B. (1996). Ionization states of the catalytic residues in HIV-1 protease. *Nature Structural & Molecular Biology, 3*(11), 946-950.

Souaille, M., & Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications, 135*(1), 40-57.

Stanton, R. V., Hartsough, D. S., & Merz Jr, K. M. (1993). Calculation of solvation free energies using a density functional/molecular dynamics coupled potential. *The Journal of Physical Chemistry, 97*(46), 11868-11870.

Sussman, F., Villaverde, M. C., L Dominguez, J., & Danielson, U. H. (2013). On the active site protonation state in aspartic proteases: implications for drug design. *Current pharmaceutical design, 19*(23), 4257-4275.

Torrie, G., & Valleau, J. (1977). Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *The Journal of chemical physics, 66*(4), 1402-1408.

Torrie, G. M., & Valleau, J. P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters, 28*(4), 578-581.

Tözsér, J., Bagossi, P., Weber, I. T., Copeland, T. D., & Oroszlan, S. (1996). Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases. *Journal of Biological Chemistry, 271*(12), 6781-6788.

Trylska, J., Bała, P., Geller, M., & Grochowski, P. (2002). Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease. *Biophysical journal, 83*(2), 794-807.

Trylska, J., Grochowski, P., & McCammon, J. A. (2004). The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. *Protein science, 13*(2), 513-528.

Tsukamoto, S., Sakae, Y., Itoh, Y., Suzuki, T., & Okamoto, Y. (2018). Computational analysis for selectivity of histone deacetylase inhibitor by replica-exchange umbrella sampling molecular dynamics simulations. *The Journal of chemical physics, 148*(12), 125102.

Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. (1985). Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science, 229*(4720), 1402-1405.

Wang, Y.-X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., . . . Torchia, D. A. (1996). Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry, 35*(31), 9945-9950.

Weber, I. T., Cavanaugh, D. S., & Harrison, R. W. (1998). Models of HIV-1 protease with peptides representing its natural substrates. *Journal of Molecular Structure: THEOCHEM, 423*(1-2), 1-12.

Wlodawer, A., & Erickson, J. W. (1993). Structure-based inhibitors of HIV-1 protease. *Annual review of biochemistry, 62*(1), 543-585.

Woodcock, H. L., Hodošček, M., & Brooks, B. R. (2007). Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms. *The Journal of Physical Chemistry A, 111*(26), 5720-5728.

Yildirim, I., Park, H., Disney, M. D., & Schatz, G. C. (2013). A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. *Journal of the American Chemical Society, 135*(9), 3528-3538.

Exploring the Concerted Mechanistic Pathway for HIV-1 PR—Substrate Revealed by Umbrella Sampling Simulation

Zainab K. Sanusi,[a] Monsurat M. Lawal,[a] Pancham Lal. Gupta,[d] Thavendran Govender,[b] Sooraj Baijnath,[a] Tricia Naicker,[a] Glenn E. M. Maguire,[a, c], Bahareh Honarparvar,[a] Adrian E. Roitberg,[d] and Hendrik G. Kruger.[a*]

[a]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

[b]AnSynth PTY LTD, 498 Grove End Drive, Durban 4001, South Africa.

[c]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.

[d]Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States

*__Corresponding authors__: Prof. H. G. Kruger; kruger@ukzn.ac.za

Table of contents

The HIV-1 PR cleavage sites in the Gag and Gag-Pol polyproteins.

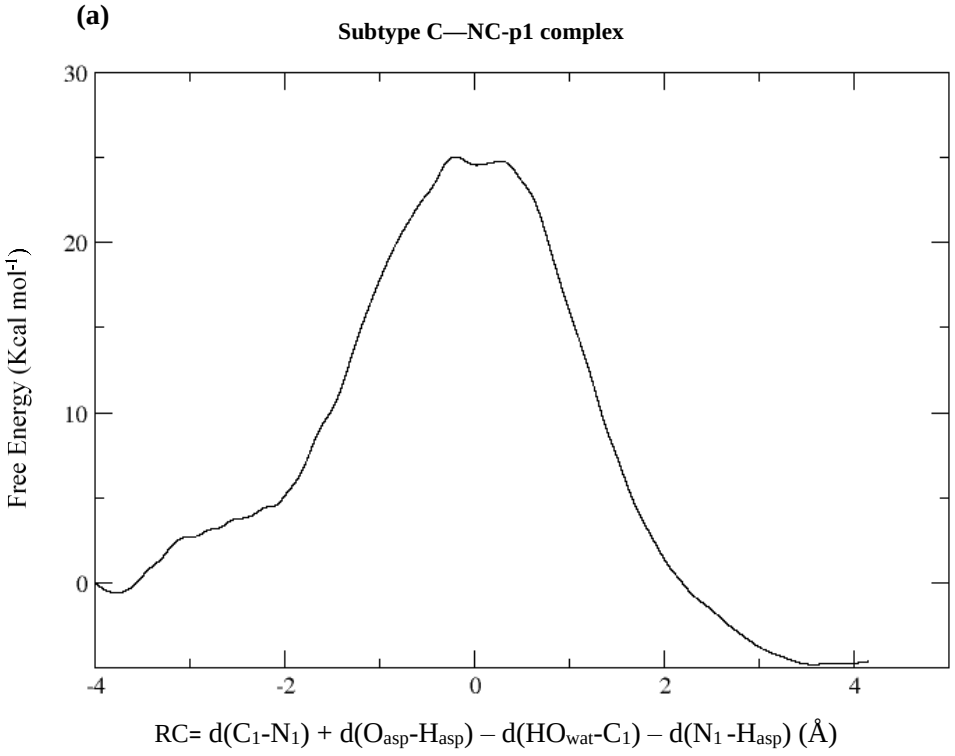
Table S**1**. Recognition sequences for natural substrates cleaved by the HIV-1 protease.[1, 2]

Peptide sequences cleavage domain										Cleavage domain
P5	P4	P3	P2	P1	*	P1′	P2′	P3′	P4′	P5′

Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln-Asn	MA-CA
Glu-Arg-Gln-Ala-Asn*Phe-Leu-Gly-Lys-Ile	NC-p1
Cys-Thr-Leu-Asn-Phe*Pro-Ile-Ser-Pro-Ile	PR-RT

The asterisk (*) denotes the scissile bond. Matrix-capsid; MA-CA, NC-p1, *trans* frame peptide-protease; PR-RT, reverse transcriptase-RNAseH. The P1, P2…P5 to P1′, P2′…P5′ represents the amino acid residues of the substrate and are counted from the cleavage point.

The free energy plot for the concerted acyclic mechanism for all considered enzyme—substrate and 3D structures of reactant (RC), transition state (TS), and product complex (PC) for the acylation step of the HIV-1 PR enzyme.



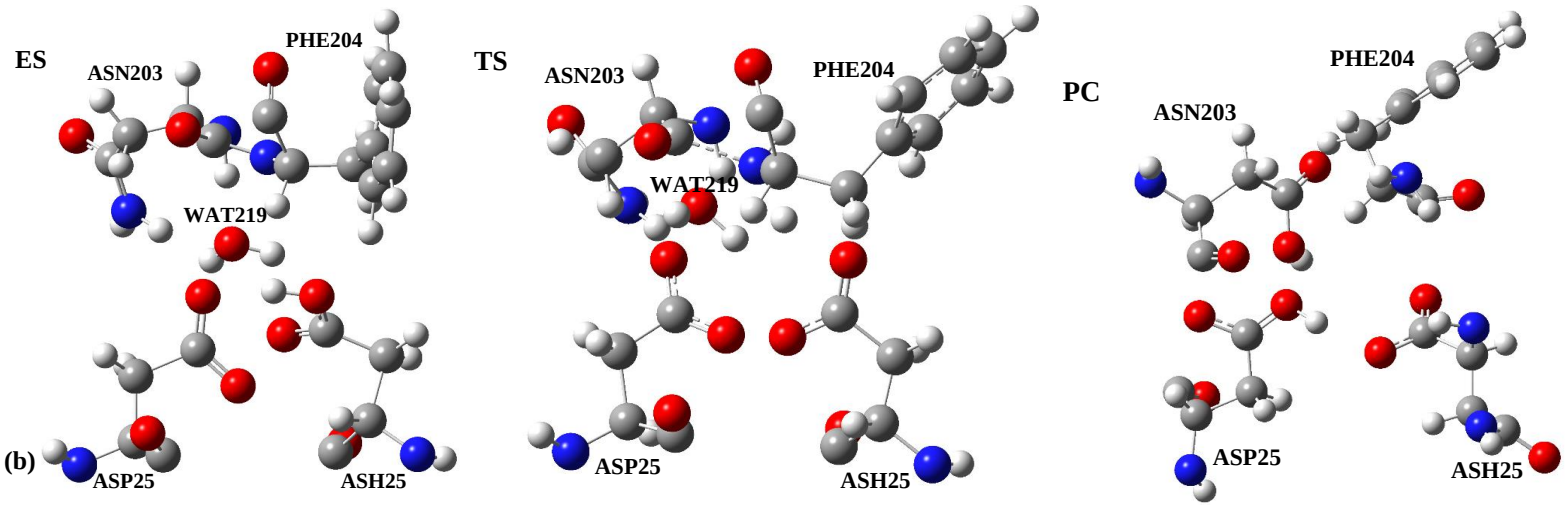
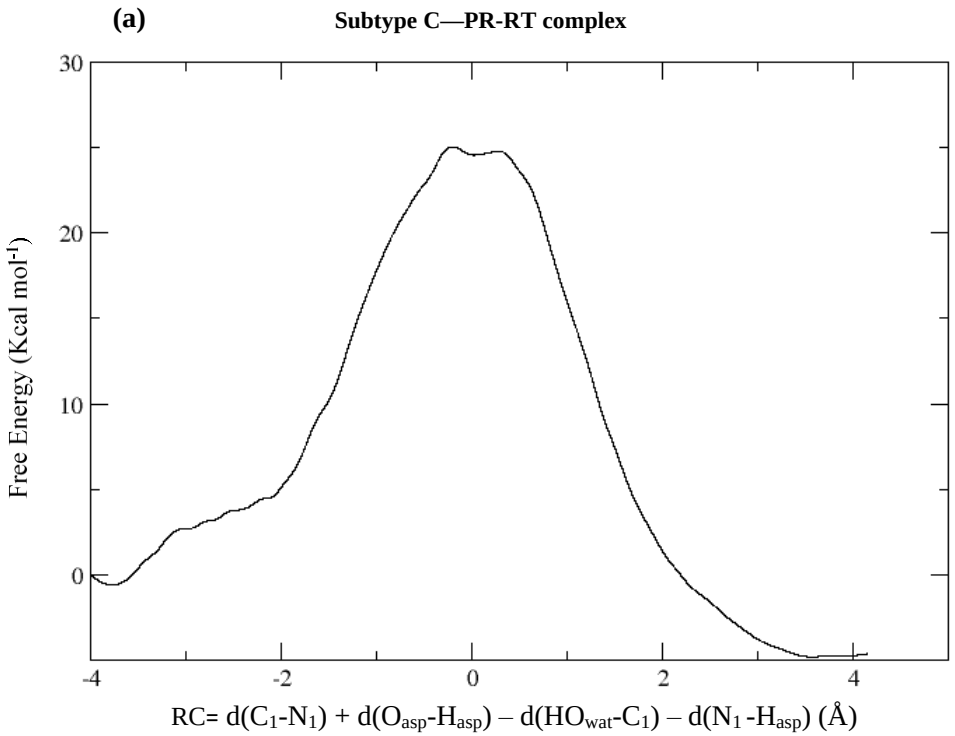


Figure S1. (a) Free energy profile for the concerted acyclic mechanism of subtype C HIV-1 PR with NC-p1 natural substrate. (b) 3D structures of enzyme—substrate (ES), transition state (TS), and product complex (PC) for the acylation step of the HIV-1 PR enzyme taken from trajectories.



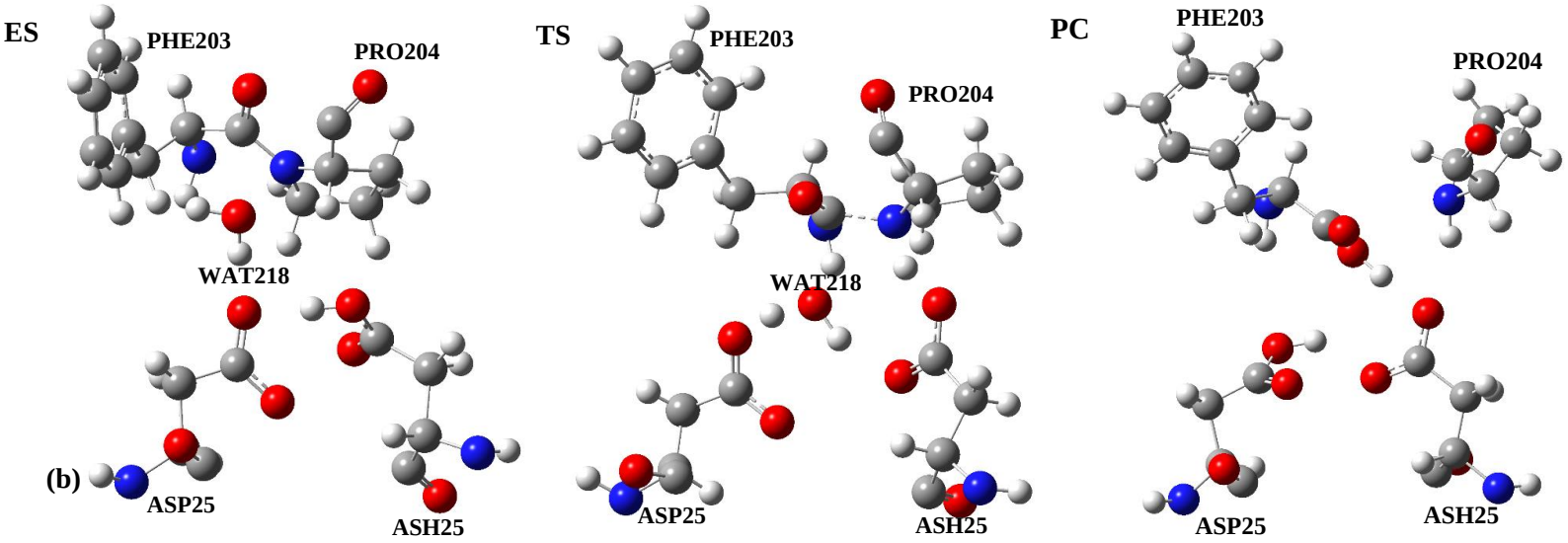


Figure S2. (a) Free energy profile for the one-step concerted acyclic mechanism of subtype C HIV-1 PR with PR-RT natural substrate. (b) 3D structures of enzyme—substrate (ES), transition state (TS), and product complex (PC) for the acylation step of the HIV-1 PR enzyme taken from trajectories.

The averaged distances for reaction complex (RC), transition states (TSs) and product complex (PC) structures for the concerted mechanism is given below.

Table S**2**. Averaged distances for RC, TS and PC structures for the concerted mechanism obtained by DFTB/MM 1D-PMF for subtype B—NC-p1complex

DFTB/MM	**RC**	**TS**	**PC**
$d(HO_{wat}\text{-}C_1)$	2.81	2.06	1.31
$d(C_1\text{-}N_1)$	1.36	1.75	3.09
$d(N_1\text{-}H_{asp})$	3.23	1.10	1.06
$d(O_{asp}\text{-}H_{asp})$	1.00	1.47	2.85

Values are reported in Å.

Table S**3**. Averaged distances for RC, TS and PC structures for the concerted mechanism obtained by DFTB/MM 1D-PMF for subtype C—NC-p1 complex

DFTB/MM	**RC**	**TS**	**PC**
$d(HO_{wat}\text{-}C_1)$	2.75	2.43	1.35
$d(C_1\text{-}N_1)$	1.32	1.93	3.88
$d(N_1\text{-}H_{asp})$	3.20	1.10	1.03
$d(O_{asp}\text{-}H_{asp})$	1.02	1.48	2.34

Values are reported in Å.

Table S**4**. Averaged distances for RC, TS and PC structures for the concerted mechanism obtained by DFTB/MM 1D-PMF for subtype C—PR-RT complex

DFTB/MM	**RC**	**TS**	**PC**
$d(HO_{wat}\text{-}C_1)$	2.98	2.20	1.42
$d(C_1\text{-}N_1)$	1.42	1.62	3.28
$d(N_1\text{-}H_{asp})$	3.52	1.20	1.02
$d(O_{asp}\text{-}H_{asp})$	0.99	1.32	2.12

Values are reported in Å.

References

1. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2000) How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of HIV-1 protease1, *Journal of molecular biology 301*, 1207-1220.
2. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, *Structure 10*, 369-381.

CHAPTER EIGHT

CONCLUSION

The HIV-1 virus is among the most investigated infection globally, and the discovery of zidovudine in the 80s was the breakthrough that led to the development of many other antiretroviral drugs. These antiretroviral drugs inhibit certain developmental phase/enzymes in the life circle of the human immune virus and have remained a well-known method of therapeutic treatment, as a result, HIV-1 PR is still a therapeutic target enzyme. The mechanism of action of proteases on the cleavage of the natural polypeptides has been a focus of research for the past decades. Despite important contributions on understanding the recognition and reaction mechanism of the HIV-1 PR with respect to its natural substrates across the globe, some important gaps in knowledge are yet to be addressed by researchers.

The introductory chapter focuses on the main topic which provides a background discourse necessary for the rest of the thesis. The different theoretical techniques accessible, computational methods and tools used are highlighted in chapter two. Presented in chapter three is a literature review on the recognition and hydrolysis of the natural substrate by HIV-1 PR. The mechanism of hydrolysis of the natural substrates by the HIV-1 PR enzyme at the molecular level (using an enzyme—substrate system and a hybrid QM/MM method) was presented in chapter four. Chapter five analyzed the recognition pattern of two asymmetric natural substrates in the scissile and non-scissile bond by homodimeric subtypes B and C-SA HIV-1 PR using a hybrid QM/MM (ONIOM) method. This study was then broadened (Chapter six) to understand the cleavage mechanism of all the nine asymmetric natural substrates by subtypes B and C-SA HIV-1 PR with the application of a hybrid QM/MM (ONIOM) technique. Finally, umbrella sampling was employed to ascertain our proposed model in the recognition pattern and hydrolysis of the natural substrates by HIV-1 PR in chapter seven.

The HIV-1 PR reaction mechanism has frequently been studied as a stepwise and/or general acid-base process rather than a concerted reaction. However, debates such as the exact rate-limiting step seems to occur with the stepwise reaction process of HIV-1 PR—substrate. Likewise, the protonation pattern of the catalytic aspartates at the active site which is responsible for the

hydrolytic cleavage of the natural substrates remains a topic of discussion and was addressed in the literature review (Chapter three).

An identifiable gap from the literature review incited the design of a concerted six-membered general acid-base mechanism of HIV-1 PR—substrate model, which occurs through cyclic transition state (TS) as proposed from experiment. Based on previous theoretically investigations for similar six-membered ring transition state structures, a concerted chemical process was studied with the aim of providing alternative insight into the catalytic mechanism of HIV-1 PR and inhibitor design through computational methods. It was observed that the mechanism that provides a marginally lower activation energy involves an acyclic TS model (Chapter four) with one water molecule at the HIV-1 PR active site.

A follow up investigation using the proposed concerted acyclic TS modeling *via* a hybrid QM/MM ONIOM approach to account correctly for the scissile bond cleavage (recognition pattern) of the natural substrates by subtypes B and C-SA HIV-1 PRs was presented in chapter five. It was observed that optimum recognition and specificity depends on the length of the amino acids sequence as well as the structural detail of the substrates. Both subtypes B and C-SA HIV-1 PR recognize and cleave at the scissile and non-scissile bond of the natural substrate sequence with a lower activation free energy than the non-scissile region. Subsequent investigation (Chapter six) was dedicated to a more broadened study of all the remaining natural substrates as described earlier, using a longer amino acid sequence (P5-P5′). It was observed that one of the natural substrates (NC-p1) has a characteristic slower rate of hydrolysis and is the rate-limiting cleavage step in the maturation of the Gag and Gag-pol substrate, this was observed in experimental work also. It was concluded that our computational model also provides excellent results for biologically active mutants of HIV-1 PR in that it correctly recognizes and cleaves its natural substrate polypeptides.

Umbrella sampling was explored by applying a QM/MM MD method to investigate and ascertain our proposed theoretical model for substrate recognition and reaction mechanism of mainly the C-SA HIV-1 PR (Chapter seven). The theoretical findings suggest that the proposed concerted acyclic TS modeling corresponds in principle with experimental data with lower activation energies. This method complimented our previous mechanistic studies of the HIV-1 PR, presenting the overall free energy profile of the HIV-1 PR potential energy surface.

Studies on the HIV-1 PR are quite diverse and theoretical approaches have played a vital role in determining the recognition pattern and enzymatic mechanism of HIV-1 PR and its natural target. The present work is dedicated to the study of the entire HIV-1 PR system considered and its exact Gag and Gag-pol polypeptide sequence. Tremendous improvement is expected in years to come through more advanced computer soft and hardware development, which would elucidate the general reaction pathway for the HIV-1 PR—substrate/ligand complex. This computational technique can be utilized to rationalize lead compounds against specific target. The possibility of integrated computational algorithms which do not involve restraining/partitioning/constraining/cropped model system of enzyme—substrate (ES) mechanism would likely surface in future to perfectly model the HIV-1 PR catalytic process on natural substrates/inhibitors.