UNIVERSITY OF KWAZULU-NATAL

A COMPUTATIONAL PERSPECTIVE ON THE CONCERTED CLEAVAGE MECHANISM OF THE NATURAL TARGETS OF HIV-1 PROTEASE

2018

MONSURAT MOTUNRAYO LAWAL

214584429

A COMPUTATIONAL PERSPECTIVE ON THE CONCERTED CLEAVAGE MECHANISM OF THE NATURAL TARGETS OF HIV-1 PROTEASE

MONSURAT MOTUNRAYO LAWAL

214584429

2018

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. Typically, most of the chapters would be published in internationally recognized, peer-reviewed journals.

This is to certify that the content of this thesis is the original research work of Monsurat Motunrayo Lawal.

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

Supervisor: Signed: --  --Name: **Prof. H. G. Kruger** Date: 30/10/2018

Co-Supervisor: Signed:  Name: **Dr. B. Honarparvar** Date: 30/10/2018

Co-Supervisor: Signed: ----------Name: **Dr. G. E. M. Maguire** Date: 30/10/2018

DEDICATION

This dissertation is proudly dedicated to my parents Mr. Ibrahim and Mrs. Ooreofe Lawal, for the unquantifiable values instilled in me. You two are the best.

ABSTRACT

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) caused by the Human Immunodeficiency Virus (HIV). A major breakthrough in the treatment of HIV-1 was the use of drugs inhibiting specific enzymes necessary for the replication of the virus. 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. The mechanism of action of the HIV-1 PR on the proteolysis of these polyproteins has been a subject of research over the past three decades.

Most investigations on this subject have been dedicated to exploring the reaction mechanism of HIV-1 PR on its targets as a stepwise general acid-base process with little attention on a concerted model. One of the shortcomings of the stepwise reaction pathway is the existence of more than two TS moieties, which have led to varying opinions on the exact rate-determining step of the reaction and the protonation pattern of the catalytic aspartate group at the HIV-1 PR active site. Also, there is no consensus on the actual recognition mechanism of the natural substrates by the HIV-1 PR.

By means of concerted transition state (TS) structural models, the recognition mode and the reaction mechanism of HIV-1 PR with its natural targets were investigated in this present study. The investigation was designed to elucidate the cleavage of natural substrates by HIV-1 PR using the concerted TS model through the application of computational methods to unravel the recognition and reaction process, compute activation parameters and elucidate quantum chemical properties of the system.

Quantum mechanics (QM) methods including the density functional theory (DFT) models and Hartree-Fock (HF), molecular mechanics (MM) and hybrid QM/MM were employed to provide better insight in this topic. Based on experience with concerted TS modelling, the six-membered ring TS structure was proposed. Using a small model system and QM methods (DFT and HF), the enzymatic mechanism of HIV-1 PR was studied as a general acid-base model having both catalytic aspartate group participating and water molecule attacking the natural substrate synchronously. The natural substrate scissile bond strength was also investigated via changes of electronic effects. The proposed concerted six-membered ring TS mechanism of the natural substrate within the entire enzyme was studied using hybrid QM/MM; “Our own N-layered

Integrated molecular Orbital and molecular Mechanics" (ONIOM) method. This investigation led us to a new perspective in which an acyclic concerted pathway provided a better approach to the subject than the proposed six-membered model. The natural substrate recognition pattern was therefore investigated using the concerted acyclic TS modelling to examine if HIV-1 (South Africa subtype C, C-SA and subtype B) PRs recognize their substrates in the same manner using ONIOM approach.

A major outcome in the present investigation is the computational modelling of a new, potentially active, substrate-based inhibitor through the six-membered concerted cyclic TS modelling and a small system. By modelling the entire enzyme—substrate system using a hybrid QM/MM (ONIOM) method, three different pathways were obtained. (1) A concerted acyclic TS structure, (2) a concerted six-membered cyclic TS model and (3) another six-membered ring TS model involving two water molecules. The activation free energies obtained for the first and the last pathways were in agreement with in vitro HIV-1 PR hydrolysis data. The mechanism that provides marginally the lowest activation barrier involves an acyclic TS model with one water molecule at the HIV-1 PR active site. The outcome of the study provides a plausible theoretical benchmark for the concerted enzymatic mechanism of HIV-1 PRs which could be applied to related homodimeric protease and perhaps other enzymatic processes. Applying the one-step concerted acyclic catalytic mechanism for two HIV-1 PR subtypes, the recognition phenomena of both enzyme and substrate were studied. It was observed that the studied HIV-1 PR subtypes (B and C-SA) recognize and cleave at both scissile and non-scissile regions of the natural substrate sequences and maintaining preferential specificity for the scissile bonds with characteristic lower activation free energies.

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. Theoretical investigations on the enzymatic mechanism of HIV-1 PR—natural substrate in years to come, would likely involve the application of sophisticated computational techniques aimed at exploring more than the energetics of the system. The possibility of integrated computational algorithms which do not involve partitioning/restraining/constraining/cropped model systems of the enzyme—substrate mechanism would likely surface in future to accurately elucidate the HIV-1 PR catalytic process on natural substrates/ligands.

ISIQEPHU

Isifo esisodwa esithathelwanayo esiye saba nethonya elijulile lempilo emphakathini emashumini eminyaka yamuva yi-Acquired Immune Deficiency Syndrome (i-AIDS) ebangelwa i-Human Immunodeficiency Virus (i-HIV). Ukuphumelela okukhulu ekwelapheni i-HIV-1 kwakuwukusetshenziswa kwezidakamizwa ezivimbela ama-enzyme akhethekile adingekayo ekuphindaphindiwe kwegciwane. Phakathi kwala ma-enzyme yi-protease ye-HIV-1 (PR), okuyinto enzyme ebaluleke kakhulu edingekayo ekuhlanjeni kweproteolytic ye-Gag ne-Gag-Pol polyproteins, edingekayo ekuthuthukiseni ama-virion amaprotheni avuthiwe. Indlela yokwenziwa kwe-HIV-1 PR mayelana ne-proteolysis yalezi polyproteins ibe yinto yokucwaninga kule minyaka engamashumi amathathu edlule.

Ucwaningo oluningi ngalolu daba luye lwazinikezelwa ekuhloleni indlela yokusabela ngayo i-HIV-1 PR ngezinhloso zayo njengenqubo ejwayelekile ye-asidi-base ngaphandle kokunakekelwa okuncane emfanekisweni owenziwe kahle. Enye yeziphambeko zendlela yokuphendula ngesinyathelo esithinta izinyathelo ezikhona yi-existence yama-TS miieties angaphezu kwama-2, okuye kwaholela emibonweni ehlukahlukene ngesilinganiso esinqunyiwe-esinqumayo sendlela yokusabela kanye nephethini ye-protonation yeqembu le-aspartate le-catalytic ku-HIV-1 PR indawo esebenzayo. Futhi, akukho ukuvumelanisa mayelana nokwaziswa okubonakalayo kwamagatsha emvelo nge-HIV-1 PR.

Ngendlela yokuhlelwa kwesimo sokuguquguquka (TS) esakhiweni, isimo sokuqaphela kanye nendlela yokusabela ngayo i-HIV-1 PR kanye nezinhloso zayo zemvelo zaphenywa kulolu cwaningo. Uphenyo lwenzelwe ukucacisa ukucaciswa kwezinsizakalo zemvelo nge-HIV-1 PR usebenzisa i-model ehambisana ne-TS ngokusebenzisa izindlela zokusebenzisa ulwazi ukuze kutholakale inqubo yokuqaphela nokuphendula, ukuqhathanisa imingcele yokusebenza kanye nokwehlukanisa izici zamakhemikhali ezinqwaba zesistimu.

Izindlela ze-Quantum mechanics (QM) ezihlanganisa izinhlobo ze-theory (DFT) ze-theory (DFT) kanye ne-Hartree-Fock (HF), ama-mechanical mechanics (MM) kanye ne-hybrid QM / MM basebenziwe ukunikeza ukuqondisisa kangcono kule sihloko. Ngokususelwa ekuhlangenwe nakho ngokufometha kwe-TS ehlelekile, isakhiwo se-TS ring ring sezinyanga eziyisithupha siphakanyisiwe. Ukusebenzisa uhlelo oluthile lwamamodeli kanye nezindlela ze-QM (i-DFT ne-HF), uhlelo lwe-enzymatic lwe-HIV-1 PR lwalufundwa njengendlela ejwayelekile ye-asidi-base equkethe iqembu elilodwa le-aspartate elithatha iqhaza futhi i-molecule yamanzi ehlasela i-

substrate yemvelo ngokuvumelana. Amandla esibopho esivumelana ne-substrate abuye aphishwe ngezinguquko zemiphumela ye-elektroniki. Uhlelo oluthile oluhlelwe ngamasithandathu e-ring TS lwe-substrate yemvelo kulo lonke i-enzyme lucwaningo lwasetshenziswa nge-hybride QM / MM; "I-N-layered Integrated Yomlomo We-Orbital Yamasosha Yomzimba Nama mechanical Mechanics" (ONIOM). Lolu uphenyo lisiholela esimweni esisha lapho i-acyclic endleleni ekhonjiwe yanikeza indlela engcono kakhulu kulo mbandela kunomfanekiso ohlongozwayo onamathandathu. Ngakho-ke i-substrate recognition pattern iphethwe ngokucubungula ngokusebenzisa i-acyclic TS modeling ehlolwe ukuhlola ukuthi ngabe i-HIV-1 (i-South Africa subtype C, C-SA kanye ne-subtype B) i-PRs yaqaphela ama-substrate awo ngendlela efanayo ngokusebenzisa i-ONIOM.

Umphumela omkhulu kulo uphenyo lwamanje yi-modeling ye-computational ye-inhibitor entsha, engase isebenze, esekelwe ngaphansi kwe-sub-based ngokusebenzisa i-TS modeling ye-cyclic esebenzayo enezinhlobo eziyisithupha kanye nesistimu encane. Ngokufanekisela yonke uhlelo lwe-enzyme-substrate esebenzisa indlela ye-QM / MM (i-ONIOM) ye-hybrid, kutholakale izindlela ezintathu ezahlukene. (1) Isakhiwo se-TS acyclic ehambisanayo, (2) imodeli ye-cyclic TS enezinhlobo eziyisithupha kanye (3) nesinye isakhiwo se-TS esinezinhlobo eziyisithupha esihlanganisa ama-molecule amabili wamanzi. Amandla mahhala okuqalisa ukusebenza atholakele ezindleleni zokuqala nezokugcina avumelana ne-in vitro HIV-1 PR ye-hydrolysis data. Indlela ehlinzekela ngokuyisisekelo isivinini sokusebenza esincane kunazo zonke kuhilela imodeli ye-acyclic TS ngenye yamakhemikhali yamanzi kwisayithi esebenzayo le-HIV-1 PR. Imiphumela yocwaningo inikeza uphawu lokukhomba olucacile lwe-enzymatic mechanism ye-HIV-1 PRs engasetshenziswa kuma-prodiase e-homodimeric kanye namanye ama-enzymatic processes. Ukusebenzisa isinyathelo esisodwa se-acyclic catalytic ehambisana ne-HIV-1 PR subtypes, izinto eziqashelwa kokubili i-enzyme kanye ne-substrate zafundwa. Kuye kwaqaphela ukuthi abafundile i-HIV-1 PR subtypes (B no-C-SA) bayaqaphela futhi banamathele kuzo zonke izifunda ezingenakusiza futhi ezingenakusiza zokulandelana kwemvelo yemvelo nokugcina imininingwane ekhethekileyo yezibopho ezikhanyelayo ezinezici ezincane zokuqalisa ukusebenza.

Ucwaningo lwesikhathi esizayo mayelana nendlela yokuphendula nge-HIV-1 PR ne-substrates zemvelo kufanele kuhileleke ekusetshenzisweni kwamasu okucubungula okuphambili ekuhlinzekeni izimpendulo ezibonakalayo kwezinye izingqinamba ezingaxazululwa. Uphenyo

lwama-theory mayelana ne-enzymatic mechanism ye-HIV-1 PR-nature substrate eminyakeni ezayo, kungenzeka ukuthi ihileleke ukusetshenziswa kwezinkambiso ezinkimbinkimbi zokucubungula ezihloswe ekuhloleni okungaphezu kwe-energetics yesistimu. I possibi

DECLARATION

I, Monsurat Motunrayo Lawal, declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original work.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

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

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

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

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

Monsurat Motunrayo Lawal

Signed: 

Date: 30/10/2018

LIST OF PUBLICATIONS

Publication one:

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.

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

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

Hendrik G. Kruger: **Supervisor**

Accepted in Current Medicinal Chemistry on 29 October 2018.

Publication two:

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

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.

Publication three:

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, she participated in the design, calculations and writing of the manuscript. Zainab K. Sanusi and Gideon F. Tolufashe 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**

Accepted in Structural Chemistry on 23 November 2018.

Publication four:

Understanding the preferential cleavage mechanism of natural substrates by HIV-1 protease subtypes

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 while **Monsurat M. Lawal** contributed in calculations and writing part of manuscript.

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

Hendrik G. Kruger: **Supervisor**

The manuscript is under preparation.

Conference contributions

- 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 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*

- The enzymatic mechanism of Human Immunodeficiency Virus type 1 protease (HIV-1 PR): are we there yet?

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

  Oral presentation at 11th CHPC National Meeting *Pretoria, SA. Dec. 4 – 8, 2017*

- The protonation state of the active site of HIV-1 protease substrate complexes: a look through ONIOM approach

  Authors: **Monsurat M. Lawal**, Bahareh Honarparvar and Hendrik G. Kruger

  Poster presented at the College of Health Science research symposium

  Durban, SA. Oct. 5 – 6 2017

ACKNOWLEDGEMENTS

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. Bahareh Honarparvar and Dr. Glenn E. M. Maguire, who adroitly supervised the entire research work. I also appreciate Zainab Sanusi, 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. Ibrahim Olufisayo and Mrs. Taibat Ooreofe Lawal, my hubby; Yakub Oluwagbemiga Seriki, my siblings; Ishaq Lawal, Basirat Alawode, Kafayat Bashir and Mosh'd Lawal, my mum-in-law; Mrs Kauthar Seriki for their love, financial, moral and spiritual support. I love you so much and God bless you all.

My appreciation 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: Steve Ibijola and Samson Akpotu for being worthy brothers, Zainab Sanusi, Rofiatu Azumah, and the Nigerian Muslims around Westville for their friendliness, love and care.

My profound gratitude also goes to these distinct persons, Bashir Alawode, Akeem O. Bashir, Imrana-B Kike, Fatima Ifeacho, Aishat Akinyemi, Aanuoluwa Babalola, Maryam Abdullahi, Adewale Adediran family, and Bode Ojugbele family, for their love, motivation, well-wishes and care.

Again, I say a big thanks to my best Boda Ayo and Gert for being more than brother and supervisor, respectively, during my postgraduate study in South Africa. Thumbs up to my hubby for putting up with my excesses during my PhD study, you are most loved, Oba. Kisses and hugs to my baby Yusra Ikeoluwa for being the calmness of my eyes, I pray you remain adorable forever.

I appreciate the 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;

Supporting information for Chapter two

Chapter three;

Supporting information for Chapter three

Chapter four;

Supporting information for Chapter four

Chapter five;

LIST OF TABLES

LIST OF ABBREVIATIONS

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

Molecular mechanics Poisson-Boltzmann surface area	MM-PBSA
Moller-Plesset	MP
Multidrug-resistant	MDR
Nuclear magnetic resonance	NMR
Nucleocapsid	NC
“Our own N-layered Integrated molecular Orbital and molecular Mechanics”	ONIOM
Particle Mesh Ewald	PME
Plasmepsin II	Plm II
Potential energy surface	PES
Protease	PR
Protein Data Bank	PDB
Quantum mechanics	QM
Replica-exchange Molecular Dynamics	REMD
Reverse transcriptase	RT
RNAse H	RH
South Africa	SA
Spacer peptide 1	p1
Spacer peptide 2	p2
Substrate-groove	S-groove
Supporting Information	SI
Thermodynamic integration	TI
Three-dimensional	3-D
Time-dependent density functional theory	TD-DFT
Transition state	TS
Tuberculosis	TB
Universal Force Field	UFF
van der Waals	VDW
Wild type	WT

TABLE OF CONTENTS

CHAPTER ONE
INTRODUCTION

1 Background

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) caused by Human Immunodeficiency Virus (HIV). It remains one of the severest health challenges.[1] Just as the name implies, HIV attacks its host's immune system thus making such individuals vulnerable to any infection. The virus has two strains; HIV-1 and HIV-2 in which the former is the most prevalent and virulent across the world while HIV-2 is mostly found in West Africa.[2] Since its identification in the 1950/60s,[3] HIV/AIDS is now a global issue with approximately 37 million persons living with the virus/disease at the end of 2016.[1] In South Africa, an average of 7.1 million persons are living with HIV/AIDS in which the province of KwaZulu-Natal has the highest prevalence.[4] The quest for managing and treating this infection is therefore a worldwide priority.

A major breakthrough in the treatment of HIV-1 is the use of drugs inhibiting specific enzymes (Figure 1) necessary for the replication of the virus.[5] 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.[6-9] The hydrolytic action of the PR on these asymmetric natural polyprotein substrate sequences results in the processing of the corresponding mature virion: Gag proteolysis provides substrates required for the processing of capsid (CA), matrix (MA), spacer peptide 1 (p1), nucleocapsid (NC), spacer peptide 2 (p2), and p6gag[10] and Gag-Pol proteolysis for reverse transcriptase (RT), RNAse H (RH), and integrase (IN).[11] The mechanism of action of the PR on the proteolysis of these polyproteins has been a subject of research over the past three decades.

Herein, the recognition and reaction mechanism of the natural Gag and Gag-Pol polyproteins segments by native and two HIV-1 PR were studied. This was carried out using theoretical tools in which molecular mechanics (MM), quantum mechanics (QM) and hybrid QM/MM methods were harnessed to solve some pertinent questions on the recognition of substrate and enzymatic catalysis of HIV-1 PR. The two major subjects (HIV-1 PR and substrate) of the research are briefly introduced and theoretical perspectives are summarized in this chapter.

2 The HIV-1 protease

The role of PR is crucial in the life cycle of HIV as this enzyme, as stated previously, catalyses the maturation of the proteins encoded in the Gag and Gag-pol gene polyprotein, thus serving as an indispensable drug target[6-9] (see **10**; Figure 1).

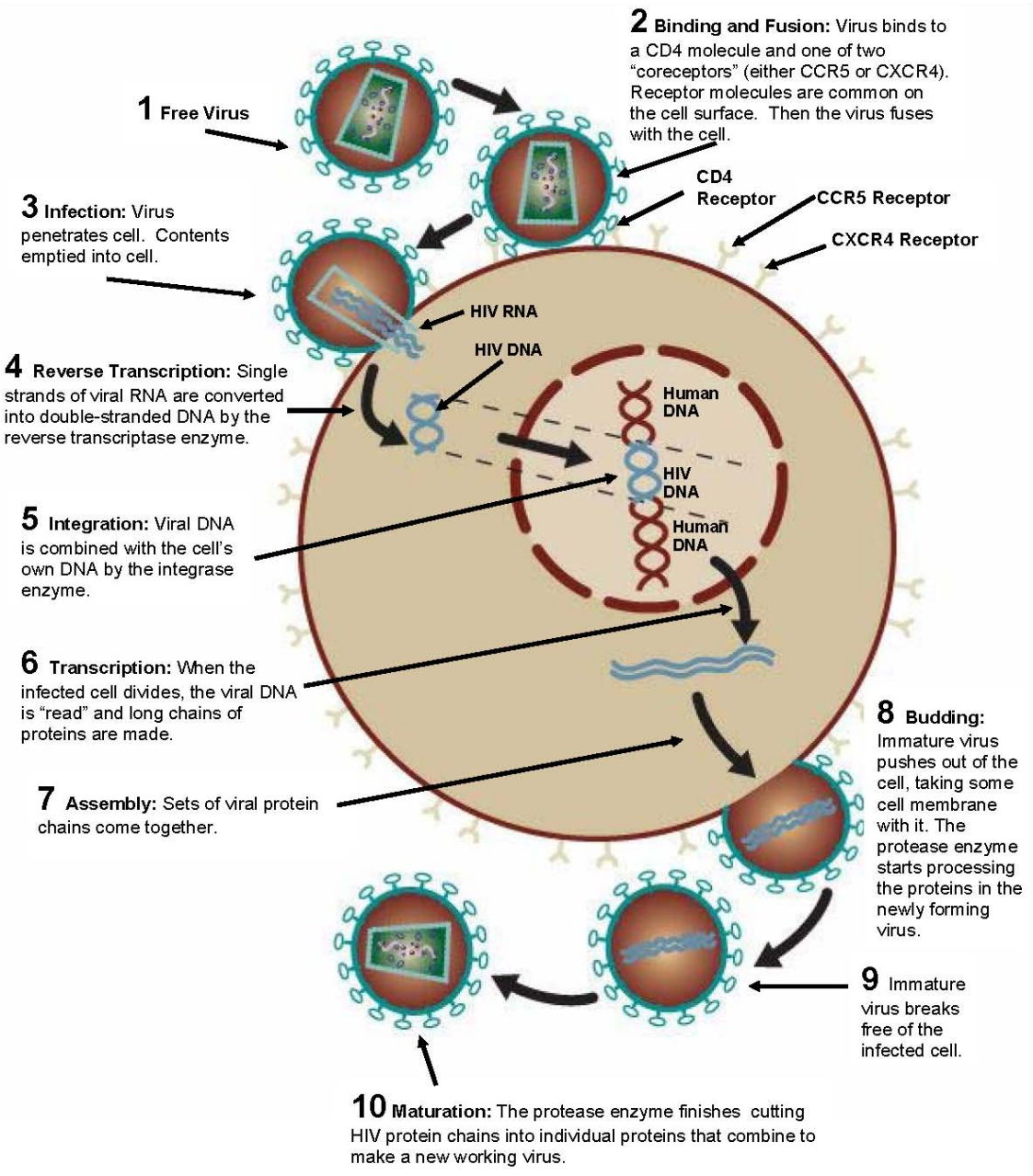


Figure 1: The HIV life cycle, copyright © 2016 International Association of Providers of AIDS Care.[12]

HIV-1 PR is a typical lentivirus proteinase, which is homodimeric with (normally) a total number of 99 amino acid residues in each monomer (Figure 2). However, variants of 100 (I36T↑T)[13] and 101 (L38L↑N↑L)[14] amino acid residues were recently reported by our group.

The two monomers of the PR are joined together to form the C_2-symmetric active homodimer[15] both contributing to the functionality of the enzyme. The enzyme has two aspartate moieties at its active site, which has long been linked to enzymatic catalysis.

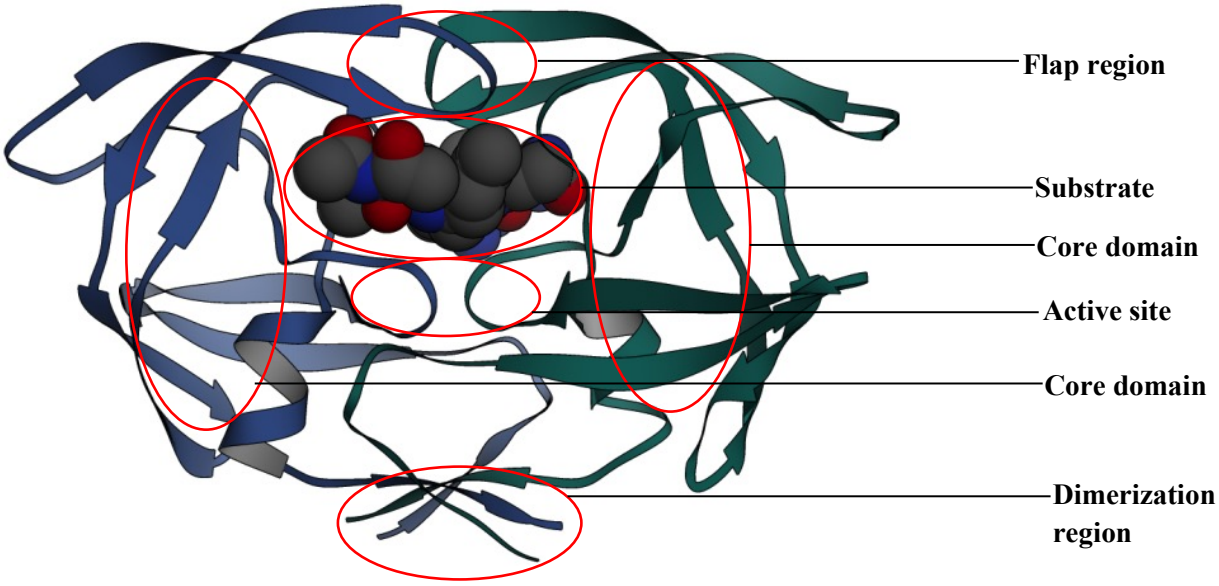


Figure 2: Structure of HIV-1 protease complexed with a natural substrate segment CA-p2; PDB code 1F7A.[16] Image created with UCSF Chimera.[17]

2.1 HIV-1 protease regions

There are three domains/regions in the HIV-1 PR namely; dimerization, core and flap domains (Figure 2). The dimerization region which is also referred to as the terminal domain, consists of the termini of four-stranded β-sheet residues 1-4, and 95-99 of each monomer, a turn encompassing residues 4-9, and the helix residues 86-94 of each monomer. This domain is crucial in dimer formation and stabilization of the HIV-1 PR.[18, 19]

The core domains are primarily composed of four-strand β structures. The domain sequence is made up of quite compact residues 10-32 and 63-85 from each monomer. The conserved Asp25-Thr26-Gly27 catalytic triad is situated at the interface of the core domains from the two monomers. The interface between the core and terminal domains is composed primarily of small hydrophobic residues. The helix of the terminal domain packs against several β strands of the core domain.[18, 19]

The flap region is composed of the most solvent exposed loop residues 33-43 preceding the β hairpin containing the flap residues 44-63. The flexible flap encloses the active site and it provides important ligand binding interactions.[18, 19]

3 The natural substrate

The hydrolytic cleavage of certain segments in the natural Gag and Gag-Pol polyprotein precursors by HIV-1 PR (Figure 3) leads to the production of specific polypeptide units, which are subsequently processed, into separated peptide units. Nine cleavage sites of these protein precursors are recognised and cleaved by HIV-1 PR to produce the corresponding peptide segments of the natural substrates (Table 1).[16, 20, 21] Inhibition of this mechanism of substrate hydrolysis by this PR has been the premise of inhibitor design[22-24] for the treatment of HIV/AIDS for many years.[5, 25] The interaction of HIV-1 PR and its natural substrates is the main research outlook of the present study, this was therefore exclusively discussed in subsequent chapters.

Table 1: The nine recognition non-homologous natural substrate polypeptides segments cleaved by the HIV-1 protease.[16, 20]

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.[16, 20]

3.1 Cleavage points

Provided in Figure 3 are the cleavage domains within the Gag and Gag-Pol gene which are recognised by the HIV-1 PR. The ease of natural substrate recognition and specificity by HIV-1 PR has been associated with substrate modulation, conserved substrate shape and interdependence conformational adaptability of both the PR and substrate.[16, 20, 26-31] A

contrasting opinion is that substrate recognition by HIV-1 PR seems to be based on the conformational specificity of the protease for the Gag and Gag-Pol polyprotein precursors.[32] The non-homologous peptide sequences of each cleavage domain have a characteristic scissile bond (asterisk in Table 1) 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 inhibitors design for this enzyme, most importantly the Phe*Pro scissile units (Table 1).[33]

In order to better examine the recognition pattern of the substrate by HIV-1 PR, a transition state modelling approach was used (Chapter five) to explore the activation barrier at different peptide bond regions within the natural substrate sequences. This enabled a comparative analysis of substrate shape specificity by the enzyme.

4 The catalytic mechanism of HIV-1 PR

The mechanistic interaction of HIV-1 PR with its natural targets is one of the most studied enzyme-substrate representatives of hydrolytic systems.[34] This hydrolysis has been widely investigated as a stepwise general acid-base mechanism involving a protonated and an unprotonated aspartate group of the PR active sites and a water molecule.[21] A variance of this mechanism is the nucleophilic model in which studies were undertaken whereby the water molecule is not involved in the calculations.[35] The latter approach seems to be diminishing as researchers continued to unravel the crucial importance of water in HIV-1 PR catalysis, as seen through crystallographic studies.[36-39]

For the present investigation, a perspective involving the molecules in a concerted general acid-base mechanism was studied (Chapters three to five). The enzymatic mechanistic pathway of HIV-1 PR—natural substrate was studied as a concerted process whereby the nucleophilic attack of water and general acid-base contributions from the aspartate group occurs concurrently. Jaskólski *et al*.[40] proposed this mechanism experimentally in 1991 and only a few studies[34, 35] have attempted to unravel the mechanism computationally with little progress. This concerted mechanistic model has been studied for a number of chemical and enzymatic reactions in our group.[41-50] We therefore applied this transition state model to describe the HIV-1 PR—natural substrate catalytic interaction using different computational approaches, which are discussed in later chapters (three to five).

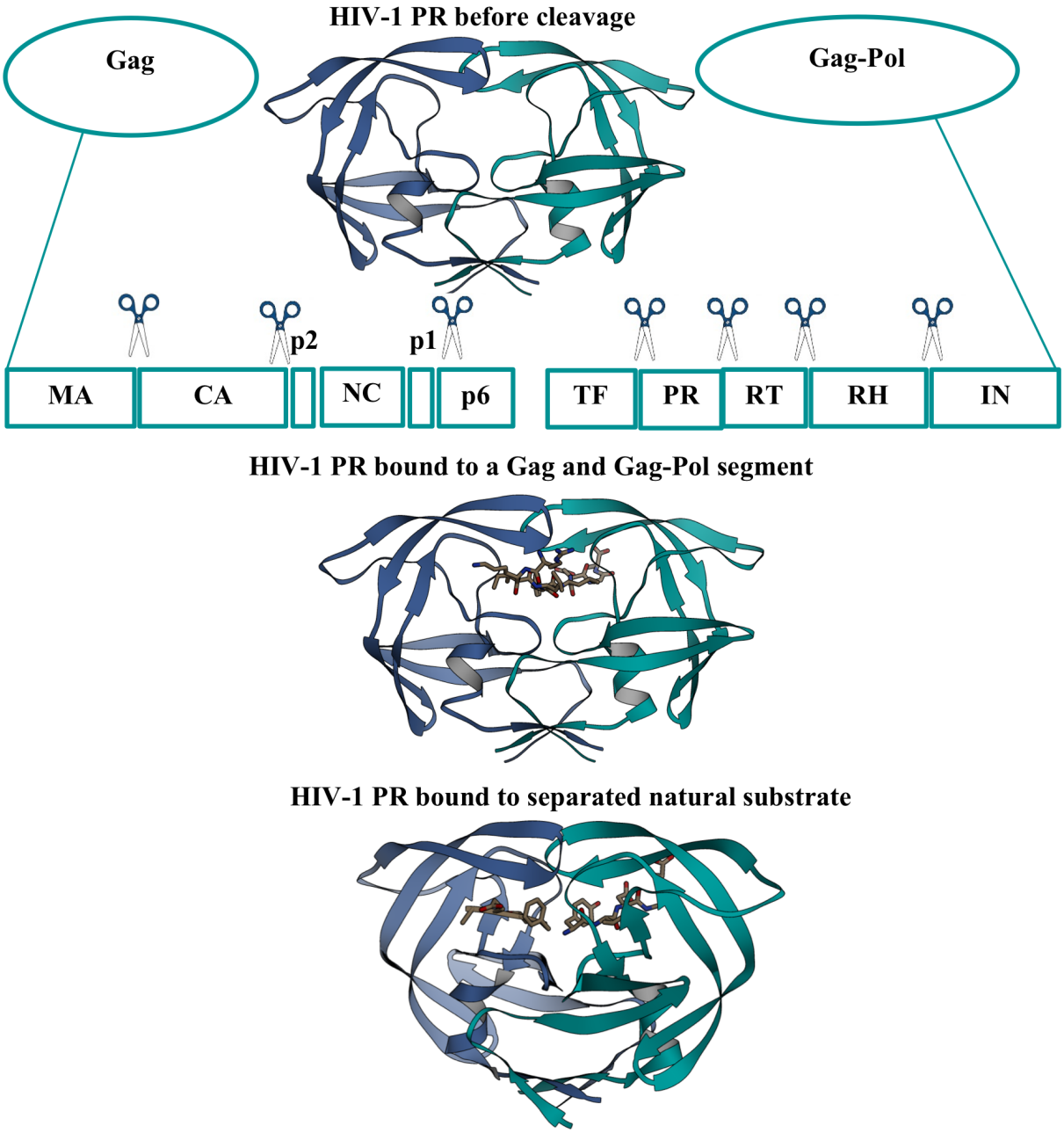


Figure 3: Recognition and reaction mechanism of HIV-1 PR, each system is based on the PDB structure indicated: the free HIV-1 PR; 1LV1,[39] substrate-bound HIV-1 PR; 1KJH[20] and broken scissile bond with the product peptides bound to HIV-1 PR; 2NPH.[51]

5 N-substituted scissile bond

For the present study, it was proposed to investigate if the strength of the scissile bond can be modified *via* electronics. Our group has some experience with N-methylation of amide bonds.[45, 52-54] The methyl group is electron donating and is expected to strengthen the amide bond through a positive inductive effect, leading to increased double bond character. Selective N-

methylation of amide nitrogen atoms increases the proteolytic stability/bioavailability.[45, 53, 55] On the other hand, electron withdrawing group such as fluorine may cause the opposite effect.

The importance of fluorine atoms in drug design has long been established and remains an ongoing perspective.[56] Fluorinated amides can be synthesized under mild reaction conditions by combining a fluoropolyalkylether substituted with one or more ester end-groups with an amine.[57]

In our research unit, peptide syntheses and applications as drug candidates has been a major outlook, most importantly for HIV-1 and tuberculosis (TB).[48] The present United States Food and Drug Administration (FDA) approved inhibitors of HIV-1 PR have been reported to be less potent for the South African HIV strain (HIV-1 C-SA),[14, 58, 59] therefore, research for better ligand binders targeting the protease enzyme is important. These commercial inhibitors bind competitively (in place of natural substrate) to the active site of the HIV-1 PR to prevent the maturation of the virus (Figure 1).

A theoretical approach to better understand the nature of factors affecting the strength of the scissile bonds in natural substrates may be useful for the design of new HIV PR inhibitors. For this study, five N-substituted substrates were computationally investigated (Chapter three) based on the same mechanism of hydrolysis of the natural substrate. The theoretical analysis of the interaction of fluorinated, methylated and other methyl substituents with HIV-1 PR was carried out to evaluate the effect of such substitution specifically at the N terminal of the scissile amide unit.

6 Computational chemistry

Computational chemistry covers a wide range of theoretical fields including, molecular mechanics and dynamics, minimization of energy, quantum mechanics and conformational analysis. These approaches and other computer-based models are used to determine the behaviour of molecular systems.[50, 60] Computational methods have been coupled with: spectroscopic, chromatographic, and other experimental methods to study drug action and inhibition; such as anticancer,[61-64] antibiotic,[65-67] antiretroviral,[48, 68] antimalarial,[69-71] and antituberculosis[72] drugs. A review paper was published in our laboratory and appeared recently, which encourages scientists to integrate experimental observations with computational methods in the drug discovery process.[48] Many studies on the inhibition of pathogens, particularly anti-HIV, antituberculosis and antiretroviral drugs have been adroitly addressed in our research unit.[48]

Chemical processes can be identified computationally through the use of advanced software parameterized to solve challenges *via* quantum mechanical (QM) or hybrid QM/MM methods. Reaction mechanisms that have not been established by experimental kinetic studies can be verified through a series of electronic structure calculations. These QM and QM/MM calculations are invaluable tools, which can be utilized to determine the feasibility of chemical reactions especially when, more than one mechanism or reaction pathway seems to be possible.[48, 60, 73] Recent reviews on various computational approaches employed by chemists to solve problems are available in literature,[48, 74-80] most of which have focussed on industrial, medicinal and chemical applications amongst others. An abstract on the choice of theoretical and computational models for this study is presented. In 2011, Pei-feng *et al*.[81] gave a review on recent advances of theoretical and computational chemistry.

6.1 Theoretical models

Models used in describing molecular systems through a particular set of approximations are referred to as theoretical models.[73, 82] Algorithms depending on approximations are then applied to atomic orbitals to calculate energies, compute frequencies and perform geometry optimizations of molecules.[73, 82, 83] Computational approaches employed in chemistry can be generalized as non-quantum mechanical and quantum mechanical methods. The non-quantum mechanical methods include molecular mechanics (MM) which in each approach is characterized by a particular force field.[82] Quantum Mechanics (QM) methods, are also known as electronic structure theories, aimed at solving the Schrödinger equation (1926)[84] to study properties of molecules.[82] A summary of required theoretical algorithm for the present investigation is provided.

6.1.1 Molecular mechanics (MM) methods

Approximations of atoms in MM simulations are done by applying classical physics laws, such as the equations of motion, to predict structures and properties of a system or a molecule. The several MM methods are characterized by their particular force field.[82] Very large systems can be modelled, including enzymes, proteins and other biological molecules. Examples of force fields are the Universal Force Field (UFF),[85] AMBER[86] and Merk Molecular Force Field (MMFF).[87] Though a less computationally expensive approximation method, MM methods cannot be applied to follow a course of reaction (bond formation/breaking) because electronic effects are neglected.[82] However, its (MM) combination with QM is now increasingly used in

practical applications, such as; the design of drug leads and catalysts, investigate enzymatic process, predict drug metabolism and resistance.[48, 74, 88]

6.1.2 *Ab initio* methods

Ab initio methods are based mainly on the laws of quantum mechanics derived strictly from theoretical principles.[82, 89, 90] Although the methods grouped under this model have the same basic approach, they vary in the mathematical algorithm approximations adopted.[91] Despite the requirements in terms of computational resources and duration, *ab initio* methods remain the most prevalent type of electronic structure method employed by computational and theoretical chemists. Compounds or systems containing up to 300 atoms can with relative ease be studied; the time required for such calculations depends on the computer hard- and software as well as the level of theory and basis set. These methods do not only provide accurately qualitative results, but also, highly quantitative information about molecular properties for a wide variety of systems.[89]

Hartree-Fock (HF), Moller-Plesset (MP*n*-including electron correlation) and Configuration Interaction (CI) are examples of *ab initio* methods. These methods systematically approach the correct "answer" as the level of theory and the size of the basis set are increased.[82] This comes at considerable cost in terms of computer resources and time. In recent years, reviews on *ab initio* calculations have been directed to specific types of calculations or studies, as this method is extensive.[92-95]

6.1.3 Density functional theory (DFT) methods

To design a more effective electronic structure method, Kohn, Sham and Hohenberg,[96, 97] proposed an alternate approach which uses density functional models. Energies are computed using electron densities instead of wave functions. DFT is also an extension of HF calculations[98] and up to 3000 atoms can be optimized with large resource requirements.[99] With computational time reduction compared to MP methods, DFT method uses approximately the same computer resources as HF theory.[83] Time-dependent density functional theory (TD-DFT) is an extension of DFT which is widely used to simulate the optical properties of both inorganic and organic compounds. An update on this model was presented by Laurent *et al*.[100] and Salahub *et al*.[101] in 2013.

Within the framework of this investigation in which Gaussian program[102] will be used, DFT methods have been categorised as: Becke three-parameter hybrid, dispersion, long-range-

corrected, Minnesota, PBE correlation, Becke one-parameter hybrid, revised B97, and τ-dependent gradient-corrected correlation functionals.[103] Examples of each functional is presented in Figure 4. The hybrid method by Becke[104] and Lee *et al.,*[105] (B3LYP) is the most popular among the DFT models (especially for organic molecules). This model uses exact exchange and gradient corrected density functional approximations to calculate correlation energies from electron densities.[104-107] Although some criticism, such as its poor estimation of barrier heights and weak interactions has been noted,[108-110] B3LYP has been embraced to give relatively good geometries of most organic and organometallic molecules.[44, 50, 111-114] The M06[108, 109] functionals are now in wide use for system modelling and calculation of molecular properties with better levels of accuracy in comparison with experiments.[46, 49] More than two DFT models were selected to investigate the aims of this present study.

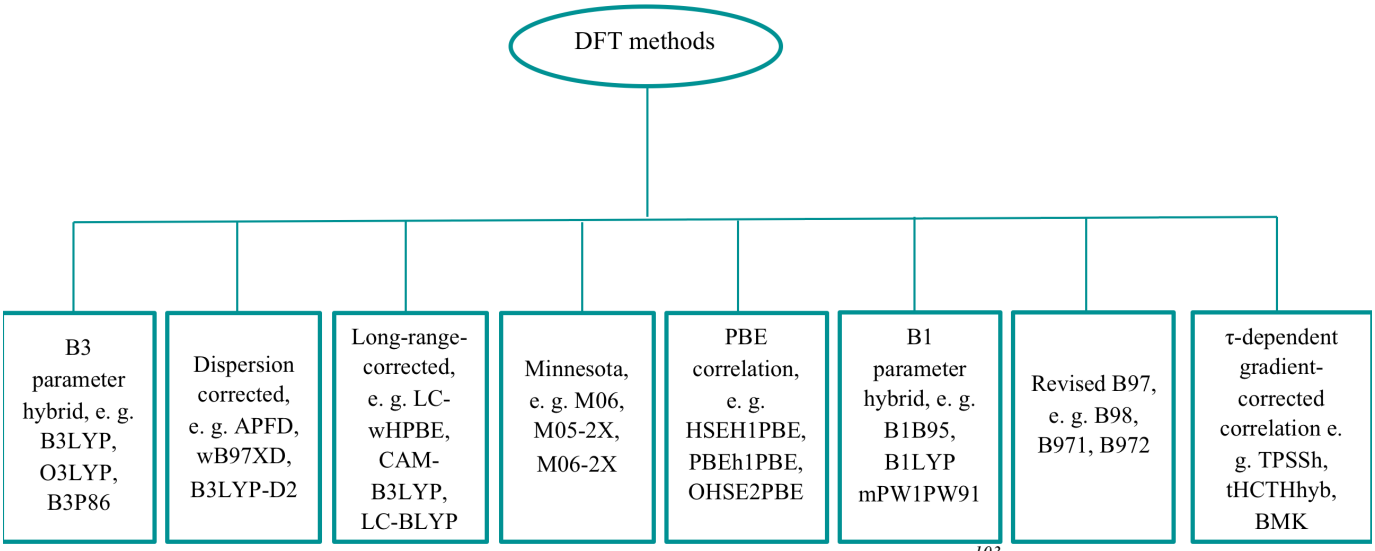


Figure 4: DFT types and examples as highlighted here.[103]

6.1.4 Hybrid QM/MM methods

In order to provide solutions to challenges in realistic theoretical modelling of large systems (such as enzymes) and taking into consideration cost effectiveness, coupling QM with MM methods was proposed.[115-117] A review on the principles and applications of these methods appear in 2012.[117] The fundamental concept of QM/MM is to divide a very large system into two or three parts based on the research interests. The QM region usually consists of the most reactive atoms that are involved in bond forming/breaking process, binding and catalysis, while the rest of the system and solvent are apportioned to the MM region.[99, 115] Based on the partitioning strategy, there exists two different forms of QM/MM of Hamiltonian known as

subtractive and additive forms.[117] "Our own N-layered Integrated molecular Orbital and molecular Mechanics" (ONIOM)[99, 118] and Umbrella sampling.[119, 120] methods are examples of subtractive and additive form, respectively. The hybrid QM/MM method offers a unique approximation in which both QM accuracy and MM efficiency is shared in molecular simulations.[117] In this present investigation, QM/MM approach was extensively used to unravel the catalytic mechanism of HIV-1 PR.

6.1.4.1 ONIOM method

A typical subtractive hybrid QM/MM method is ONIOM that allows partitioning of a molecular system into two parts namely; the model and the environment systems.[99] This method allows grouping of an entire system (such as enzyme) into layers based on the relevance of each part to the enzyme's bioactivity. Using GaussView[121] and Gaussian software,[102] three layer ONIOM segmentation could be achieved depending on the research aim. For the present study, the HIV-1 PR—substrate was subdivided into two distinct layers in which the small "model" part (QM; DFT) consists of the catalytic site of the enzyme and the natural substrate while the remaining residues were treated with MM (AMBER force field[122]) (Figure 5). This approach was recently used to estimate the binding free energies of known HIV PR inhibitors and two HIV-1 PR subtypes displaying reasonable level of accuracy when compared with experiment.[58] A review on the ONIOM method was recently done in which detailed parameterization procedure and wide range applications to varying classes of molecules and systems was discussed.[99]

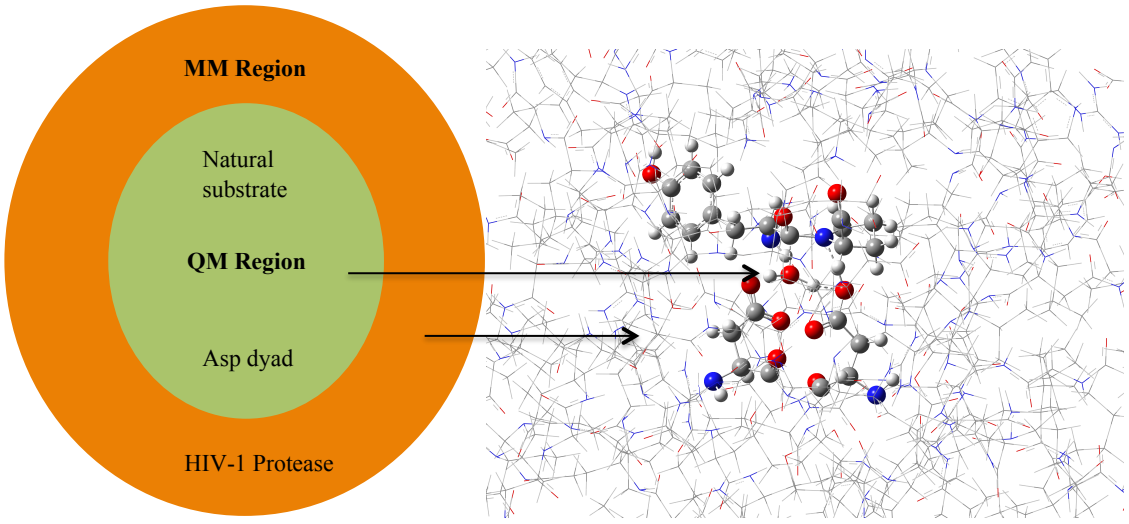


Figure 5: Schematic representation of the two-layered ONIOM model (DFT:AMBER) of HIV-1 PR—MA-CA complex.

6.1.5 Basis sets

Electronic wave functions are described mathematically with basis sets,[82] molecules are described by molecular orbitals, which are expressed in terms of atomic orbitals.[123] The larger the basis set, the better the computational output. The 6-31+G(d),[124, 125] is a renown standard basis set for calculating properties of organic molecules,[126, 127] this was applied in three parts of this present work in combination with the selected DFT methods. Other examples of basis sets are: LANL2DZ,[128-130] MidiX,[131, 132] DGDZVP,[133, 134] aug-ccpVTZ,[135-137] and def2-TZVP [138, 139] basis sets. The quality, relevance and performance of this basis sets has been studied earlier.[46] Comprehensive updates on basis sets were recently published in 2013 by Frank[126] and Grant.[127]

6.2 Computational software

The computational software required for calculations, modelling and post analyses used in this study are discussed.

6.2.1 Graphical user interfaces

GaussView,[121] UCSF Chimera,[17] PyMOL,[140] and Discovery studio[141] are the pre- and post-processor Graphical User Interfaces (GUIs) program used in the present study. All molecules were modelled, adjusted and viewed through these computational application packages.

6.2.2 The Gaussian program

Gaussian 09 Rev D.01[102] is an upgrade version of the Gaussian program which started with Gaussian 70, 92, 94, 98 and 03[142] versions. Due to the high computer resource demand for many accurate computational methods, electronic structure calculations are often carried out *via* shared processors. The enzymatic interaction of HIV-1 PR with substrates was investigated using Gaussian 09 Rev D.01[102] that is installed on clusters at the Centre for High Performance Computing in Cape Town, South Africa (www.chpc.ac.za). This package was used for all calculations including DFT and ONIOM.

6.2.3 The AMBER suite

Amongst the numerous software packages for conducting MD simulations is the AMBER[143] package, which is one of the most popular of its kind.[144] The latest version which is AMBER 17 has its roots from AMBER 4[143] which now represents 11 years of continued development (from version 9) by multiple research groups (it was 6 years[144] in 2012). AMBER program suite comprises of collections of several other programs working together to setup, perform, and

analyse MD simulations, from scratch to results analyses.[144] One of the most used versions of this software is AMBER 14[145] whose parameterization allows the use of graphics processing units (GPUs) which offers more computational power and memory bandwidth.[144] Thankfully, the Centre for High Performance Computing in Cape Town, South Africa (www.chpc.ac.za) has these extension in addition to AMBER 14[145] installed on their cluster which enabled us to obtain suitable starting structures for further calculations in this study. The latest overview of the AMBER biomolecular simulation suite appeared in 2013.[144]

7 Aims and objectives

The overall aim of this study is to investigate the recognition mechanism of natural substrate by HIV-1 PR and the subsequent catalytic interaction between them using computational methods. These are streamed down in subsequent chapters to give a full picture of the study.

Chapter One

Introductory background of study

Chapter Two

Literature review on recognition and catalysis of natural substrate by HIV-1 PR

Chapter Three

To elucidate the hydrolysis of substrates using the six-membered ring TS model (without the enzyme) as described in previous investigations[41-49] through the application of DFT methods to study the reaction process, compute activation parameters and elucidate quantum chemical properties of the system.

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 analyse variations in relation to the earlier result (small model).

Chapter Five

To investigate if the natural substrate recognition pattern is the same for HIV-1 (C-SA and subtype B) PRs.

To compare the activation energies at different peptide bond regions (other than the scissile region) within the natural substrate sequence for these enzymes using ONIOM approach.

Chapter Six

Overall conclusion of the research outcome.

8 Thesis outline

The dissertation is presented in a paper format in which each chapter is dedicated to addressing one or two research questions. In the first and the last chapters, a general introduction and an overall conclusion are provided, respectively, for the entire study. The outline is therefore highlighted.

Chapter one: General introduction of the dissertation is provided where the main direction of the study is highlighted.

Chapter two: Literature review on recognition pattern and hydrolysis of natural substrate by HIV-1 PR.

Chapter three: Modelling HIV-1 PR—substrate catalytic mechanism using small molecules and DFT methods, investigate the effects of N-methylation on the scissile bond strength.

Chapter four: Application of ONIOM to investigate substrate hydrolysis by HIV-1 PR using an exact natural substrate.

Chapter five: Substrate recognition by HIV-1 PRs (subtypes C-SA and B) using ONIOM approach, investigation of the recognition pattern and specificity of the PRs for cleavage at a particular amide bond.

Chapter six: Overall conclusion on the research outcome.

References

[1] UNAIDS. (2017) https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics.

[2] 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.

[3] Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic, *Nature 391*, 594-597.

[4] UNAIDS. (2017) http://www.unaids.org/en/regionscountries/countries/southafrica.

[5] 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*, pp 203-241, Springer.

[6] 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.

[7] 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.

[8] 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*, 280-283.

[9] 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.

[10] 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.

[11] 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.

[12] International Association of Providers of AIDS Care (2014) HIV Life Cycle, In *Fact Sheet 106* (http://www.aidsinfonet.org/fact_sheets/view/106, Ed.), AIDS infonet.

[13] Lockhat, H. A., Silva, J. R., Alves, C. N., Govender, T., Lameira, J., Maguire, G. E., Sayed, Y., and Kruger, H. G. (2016) Binding Free Energy Calculations of Nine FDA-approved Protease Inhibitors Against HIV-1 Subtype C I36T↑ T Containing 100 Amino Acids Per Monomer, *Chemical Biology & Drug Design 87, 487-498*.

[14] Maseko, S. B., Natarajan, S., Sharma, V., Bhattacharyya, N., Govender, T., Sayed, Y., Maguire, G. E., 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.

[15] Reetz, M. (1998) Preparation of novel HIV-protease inhibitors, *Chemical Communications*, 2075-2076.

[16] 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 protease, *Journal of Molecular Biology 301*, 1207-1220.
[17] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.
[18] 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.
[19] Dau, B., and Holodniy, M. (2009) Novel Targets for antiretroviral therapy, *Drugs 69*, 31-50.
[20] 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.
[21] Sussman, F., Villaverde, M. C., Dominguez, J. L., and Danielson, U. H. (2013) On the Active Site Protonation State in Aspartic Proteases: Implications for Drug Design, *Current Pharmaceutical Design 19*, 4257-4275.
[22] Makatini, M. M., Petzold, K., Alves, C. N., Arvidsson, P. I., Honarparvar, B., Govender, P., Govender, T., Kruger, H. G., Sayed, Y., Lameira, J., Maguire, G. E. M., and Soliman, M. E. S. (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors, *Journal of Enzyme Inhibition and Medicinal Chemistry 28*, 78-88.
[23] Makatini, M. M., Petzold, K., Arvidsson, P. I., Honarparvar, B., Govender, T., Maguire, G. E. M., Parboosing, R., Sayed, Y., Soliman, M. E. S., 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.
[24] Pawar, S. A., Jabgunde, A. M., Govender, P., Maguire, G. E. M., Kruger, H. G., Parboosing, R., Soliman, M. E. S., 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.
[25] Debouck, C. (1992) The HIV-1 protease as a therapeutic target for AIDS, *AIDS Research and Human Retroviruses 8*, 153-164.
[26] 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.
[27] Ozer, N., Haliloglu, T., and Schiffer, C. A. (2006) Substrate specificity in HIV-1 protease by a biased sequence search method, *Proteins: Structure, Function, and Bioinformatics 64*, 444-456.
[28] 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.
[29] Ozer, N., Schiffer, C. A., and Haliloglu, T. (2006) Predicting substrate specificity in HIV-1 protease, pp 86-87, Computer Aided Drug Design & Development Society in Turkey.
[30] Ozen, A., Haliloglu, 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.

[31] Shen, Y., Altman, M. D., Ali, A., Nalam, M. N. L., Cao, H., Rana, T. M., Schiffer, C. A., and Tidor, B. (2013) Testing the substrate-envelope hypothesis with designed pairs of compounds, *ACS Chemical Biology 8*, 2433-2441.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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*, e7860.
[38] Kumar, M., Prashar, V., Mahale, S., and Hosur, M. V. (2005) Observation of a tetrahedral reaction intermediate in the HIV-1 protease–substrate complex, *Biochemical Journal 389*, 365-371.
[39] 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.
[40] 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.
[41] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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, *Jouranl of Molecular Structure 1128*, 94-102.
[42] 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[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam, *Jouranl of Molecular Structure -Theochem 672*, 119-125.
[43] 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, *Jouranl of Molecular Structure -Theochem 771*, 165-170.
[44] 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.

[45] 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.
[46] Lawal, M. M., Govender, T., Maguire, G. E., 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.
[47] Kruger, H. G. (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *Journal of Molecular Structure: Theochem 577*, 281-285.
[48] Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design: Molecular Modeling, Spectroscopy, and Experimental Bioactivity, *Chemical Reviews 114*, 493-537.
[49] Lawal, M. M., Govender, T., Maguire, G. E., Kruger, H. G., and Honarparvar, B. (2017) 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.
[50] 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.
[51] 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, *Proceedings of the National Academy of Sciences USA 103*, 18464-18469.
[52] Zhang, S., Govender, T., Norström, T., and Arvidsson, P. I. (2005) An improved synthesis of Fmoc-N-methyl-α-amino acids, *The Journal of Organic Chemistry 70*, 6918-6920.
[53] Bose, P. P., Chatterjee, U., Hubatsch, I., Artursson, P., Govender, T., Kruger, H. G., Bergh, M., Johansson, J., and Arvidsson, P. I. (2010) In vitro ADMET and physicochemical investigations of poly-N-methylated peptides designed to inhibit Aβ aggregation, *Bioorganic & Medicinal Chemistry 18*, 5896-5902.
[54] Pratim Bose, P., Chatterjee, U., Nerelius, C., Govender, T., Norström, T., Gogoll, A., Sandegren, A., Göthelid, E., Johansson, J., and Arvidsson, P. I. (2009) Poly-N-methylated amyloid β-peptide (Aβ) C-terminal fragments reduce Aβ toxicity in vitro and in Drosophila melanogaster, *Journal of Medicinal Chemistry 52*, 8002-8009.
[55] Farahani, M. D., Honarparvar, B., Albericio, F., Maguire, G. E., Govender, T., Arvidsson, P. I., and Kruger, H. G. (2014) Proline N-oxides: modulators of the 3D conformation of linear peptides through "NO-turns", *Organic & Biomolecular Chemistry 12*, 4479-4490.
[56] Liang, T., Neumann, C. N., and Ritter, T. (2013) Introduction of Fluorine and Fluorine-Containing Functional Groups, *Angewandte Chemie International Edition 52*, 8214-8264.
[57] Falcone, S. J. (2001) Synthesis of fluorinated amides, Google Patents.
[58] Sanusi, Z. K., Govender, T., Maguire, G. E. M., Maseko, S. B., Lin, J., Kruger, H. G., and 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.
[59] Maseko, S. B., Padayachee, E., Govender, T., Sayed, Y., Kruger, G., Maguire, G. E., and Lin, J. (2017) I36T↑ T mutation in South African subtype C (C-SA) HIV-1 protease significantly alters protease-drug interactions, *Biological Chemistry 398*, 1109-1117.
[60] Schreiner, P. R. (2013) *Computational molecular science*.

[61] 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.

[62] Gupta, G., Garci, A., Murray, B. S., Dyson, P. J., Fabre, G., Trouillas, P., Giannini, F., Furrer, J., Suss-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.

[63] Bradac, 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.

[64] 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.

[65] Antony, J., Gresh, N., Olsen, L., Hemmingsen, L., Schofield, C. J., and Bauer, R. (2002) Binding of D- and L-captopril inhibitors to metallo-beta-lactamase studied by polarizable molecular mechanics and quantum mechanics, *Journal of Computational Chemistry 23*, 1281-1296.

[66] 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.

[67] Melia, C., Ferrer, S., Moliner, V., Tunon, I., and Bertran, J. (2012) Computational study on hydrolysis of cefotaxime in gas phase and in aqueous solution, *Journal of Computational Chemistry 33*, 1948-1959.

[68] 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 61*, 423-432.

[69] 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.

[70] Lacaze-Dufaure, C., Najjar, F., and Andre-Barres, 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.

[71] Mammino, L., and Bilonda, M. K. (2014) Computational study of antimalarial pyrazole alkaloids from Newbouldia laevis, *Journal of Molecular Modeling 20*, 2464.

[72] 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.

[73] Hehre, W. J. (2003) *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc 18401 Von Karman Ave., Suite 370, Irvine, CA 9261.

[74] 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.

[75] Wang, L., Nan, G., Yang, X., Peng, Q., Li, Q., and Shuai, Z. (2010) Computational methods for design of organic materials with high charge mobility, *Chemical Society Reviews 39*, 423-434.

[76] Yuriev, E., Agostino, M., and Ramsland, P. A. (2011) Challenges and advances in computational docking: 2009 in review, *Journal of Molecular Recognition 24*, 149-164.
[77] Tetko, I. V. (2012) The perspectives of computational chemistry modeling, *Journal of Computer-aided Molecular Design 26*, 135-136.
[78] van der Kamp, M. W., and Mulholland, A. J. (2013) Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology, *Biochemistry 52*, 2708-2728.
[79] Agnihotri, N. (2014) Computational studies of charge transfer in organic solar photovoltaic cells: a review, *Journal of Photochemistry and Photobiology C: Photochemistry Reviews 18*, 18-31.
[80] Deglmann, P., Schafer, A., and Lennartz, C. (2015) Application of Quantum Calculations in the Chemical Industry-An Overview, *International Journal of Quantum Chemistry 115*, 107-136.
[81] Su, P.-f., Tan, K., Wu, A.-a., Lu, X., Zhao, Y., Cao, Z.-x., and Wu, W. (2011) Advances of theoretical and computational chemistry, *Journal of Xiamen University, Natural Science 50*, 311-318.
[82] Foresman, J. B., and Frisch, A. (1996) *Exploring Chemistry with Electronic Structure Methods.*, Gaussian, Inc., Pittsburgh, PA.
[83] Tomberg, A. Gaussian 09w Tutorial: An Introduction to Computational Chemistry Using G09w and Avogadro Software, Barrett Research Group.
[84] Schrödinger, E. (1926) An undulatory theory of the mechanics of atoms and molecules, *Physical Review 28*, 1049.
[85] 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.
[86] 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.
[87] Halgren, T. A. (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, *Journal of Computational Chemistry 17*, 490-519.
[88] De Benedetti, P. G., and Fanelli, F. (2010) Computational quantum chemistry and adaptive ligand modeling in mechanistic QSAR, *Drug Discovery Today 15*, 859-866.
[89] Ghosh, S. K., and Chattaraj, P. K. *Concepts and methods in modern theoretical chemistry : electronic structure and reactivity*.
[90] Leszczynski, J. (2012) *Handbook of computational chemistry*.
[91] Lewars, E. (2011) *Computational chemistry : introduction to the theory and applications of molecular and quantum mechanics*, 2nd ed., Springer, Dordrecht Netherlands ; London ; New York.
[92] Jiang, G. L., Chen, W. J., 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*.
[93] 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, *Modeling and Simulation in Materials Science and Engineering 23*.
[94] 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*.

[95] Yanai, T., Kurashige, Y., Mizukami, W., Chalupsky, 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.
[96] Hohenberg, P., and Kohn, W. (1964) Inhomogeneous electron gas, *Physical Review 136*, B864.
[97] Kohn, W., and Sham, L. J. (1965) Self-consistent equations including exchange and correlation effects, *Physical Review 140*, A1133.
[98] Gorling, A. (1992) Kohn-Sham potentials and wave functions from electron densities, *Physical Review. A 46*, 3753-3757.
[99] 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.
[100] Laurent, A. D., and Jacquemin, D. (2013) TD-DFT benchmarks: A review, *International Journal of Quantum Chemistry 113*, 2019-2039.
[101] 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.
[102] M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox. (2013) Gaussian 09., In *Revision D.01,*, Gaussian, Inc., Wallingford CT, 2013.
[103] Gaussian. (2017) Density Functional (DFT) Methods, In *Expanding the limits of computational chemistry* (http://gaussian.com/dft/, Ed.).
[104] Becke, A. D. (1993) Densityfunctional thermochemistry. III. The role of exact exchange, *Journal of Chemical Physics 98*, 5648.
[105] Lee C., Yang W., and R.G, P. (1988) Development of Colle-Salvetti correlation-energy formula into a functional of electron density, *Phsical Review B 37*, 785-789.
[106] Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A 38*, 3098.
[107] Stephens, P., Devlin, F., Chabalowski, C., and Frisch, M. J. (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields, *The Journal of Physical Chemistry 98*, 11623-11627.
[108] 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.

[109] Zhao, Y., and Truhlar, D. G. (2008) Density functionals with broad applicability in chemistry, *Accounts of Chemical Research 41*, 157-167.
[110] 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.
[111] 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[5.4.0.0^{2,6}.03,10.0^{5,9}] undecane-8,11-lactam, *Journal of Molecular Structure - Theochem 672*, 119-125.
[112] 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.
[113] 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.
[114] W., S., Hadebe, Hendrik G. Kruger, 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.
[115] 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.
[116] Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, Wiley New York.
[117] 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.
[118] 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.
[119] 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.
[120] Kästner, J. (2011) Umbrella sampling, *Wiley Interdisciplinary Reviews: Computational Molecular Science 1*, 932-942.
[121] Dennington, R., Keith, T., and Millam, J. (2009) In GaussView, 5.0.8 ed., Semichem Inc., Shawnee Mission KS.
[122] 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.
[123] Hinchliffe, A. (1996) *Modelling molecular structures*, Wiley, New York.
[124] 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.
[125] 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.

[126] Jensen, F. (2013) Atomic orbital basis sets, *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 273-295.
[127] Hill, J. G. (2013) Gaussian basis sets for molecular applications, *International Journal of Quantum Chemistry 113*, 21-34.
[128] 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.
[129] Hay, P. J., and Wadt, W. R. (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg, *The Journal of Chemical Physics 82*, 270-283.
[130] 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.
[131] Li, J., Cramer, C. J., and Truhlar, D. G. (1998) MIDI! basis set for silicon, bromine, and iodine, *Theoretical Chemistry Accounts 99*, 192-196.
[132] 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, *Theoretical Chemistry Accounts 93*, 281-301.
[133] 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.
[134] 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.
[135] 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.
[136] 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.
[137] 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.
[138] 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.
[139] Weigend, F. (2006) Accurate Coulomb-fitting basis sets for H to Rn, *Physical Chemistry Chemical Physics 8*, 1057-1065.
[140] DeLano, W. (2002) The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA.
[141] Biovia, D. S. (2015) Discovery studio modeling environment, *San Diego: Dassault Systemes*.
[142] M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, , M. A. Robb, J. R. C., J. A. Montgomery, Jr., T. Vreven, , K. N. Kudin, J. C. B., J. M. Millam, S. S. Iyengar, J. Tomasi, , V. Barone, B. M., M. Cossi, G. Scalmani, N. Rega, , G. A. Petersson, H. N., M. Hada, M. Ehara, K. Toyota, , R. Fukuda, J. H., M. Ishida, T. Nakajima, Y. Honda, O. Kitao, , H.

Nakai, M. K., X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, , C. Adamo, J. J., R. Gomperts, R. E. Stratmann, O. Yazyev, , A. J. Austin, R. C., C. Pomelli, J. W. Ochterski, P. Y. Ayala, , K. Morokuma, G. A. V., P. Salvador, J. J. Dannenberg, , V. G. Zakrzewski, S. D., A. D. Daniels, M. C. Strain, , O. Farkas, D. K. M., A. D. Rabuck, K. Raghavachari, , J. B. Foresman, J. V. O., Q. Cui, A. G. Baboul, S. Clifford, , J. Cioslowski, B. B. S., G. Liu, A. Liashenko, P. Piskorz, , I. Komaromi, R. L. M., D. J. Fox, T. Keith, M. A. Al-Laham, , C. Y. Peng, A. N., M. Challacombe, P. M. W. Gill, , and B. Johnson, W. C., M. W. Wong, C. Gonzalez, and J. A. Pople,. (2003) Gaussian 03, , Revision B.03 ed., Gaussian, Inc. , Pittsburgh PA.

[143] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules, *Computer Physics Communications 91*, 1-41.

[144] 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.

[145] Case, D. A., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham Iii, T., Darden, T., Duke, R., and Gohlke, H. (2014) Amber 14.

CHAPTER TWO

LITERATURE REVIEW

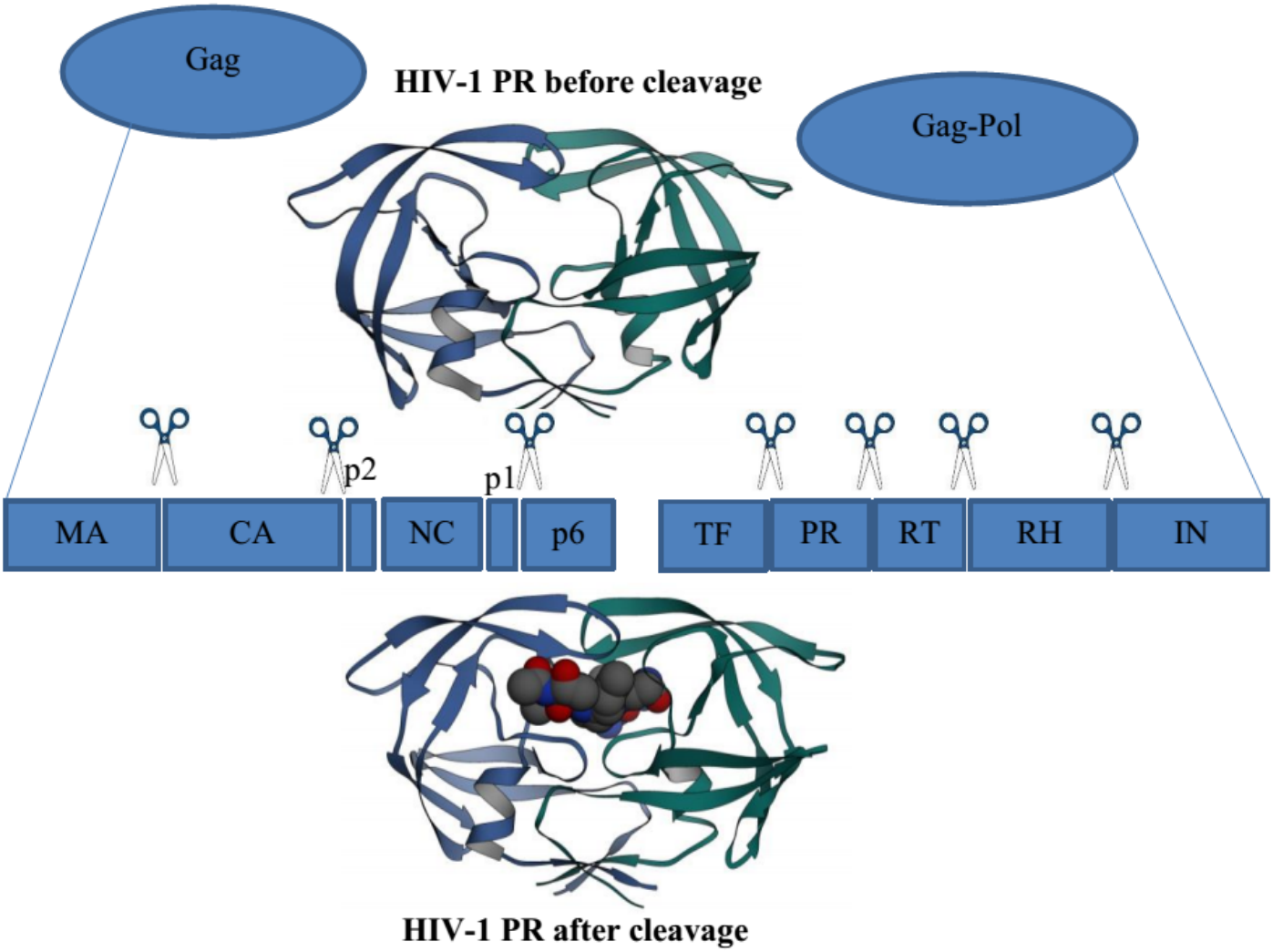
From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets

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.

Graphical Abstract



Description: Illustration of Gag and Gag-Pol cleavage process by HIV-1 PR.

Abstract

Current investigations of 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 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 is provided. This overview features more than 80 peer review articles from reputable journals. Recognition of the natural Gag and Gag-Pol cleavage junctions by this enzyme and its mutant analogues 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 better insights on the catalytic mechanism of the HIV-1 protease in the presence of its natural substrates. 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 better bioactivity.

Keywords: HIV-1 PR; Natural substrates; Recognition pattern; Reaction mechanism; Transition state modelling.

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] Notable amongst this class of enzymes are beta-secretase 1 (BACE 1), cathepsin D (Cat D), human immunodeficiency virus type 1 protease (HIV-1 PR) and plasmepsin II (Plm II). 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.[2] HIV-1 PR is an indispensable drug target moiety towards the treatment of HIV/AIDS.[3]

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)[4] and 101 (L38L↑N↑L)[5] 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.[3, 6] The hydrolytic action of the PR on these asymmetric natural polyprotein substrate sequences results in the processing of the corresponding mature virion: Gag proteolysis leads to capsid (CA), matrix (MA), spacer peptide 1 (p1), nucleocapsid (NC), spacer peptide 2 (p2) and p6gag[7]. Gag-Pol leads to reverse transcriptase (RT), RNAse H (RH), and integrase (IN).[8] The mechanism of action of the 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 (Figure 1 and Table 1).

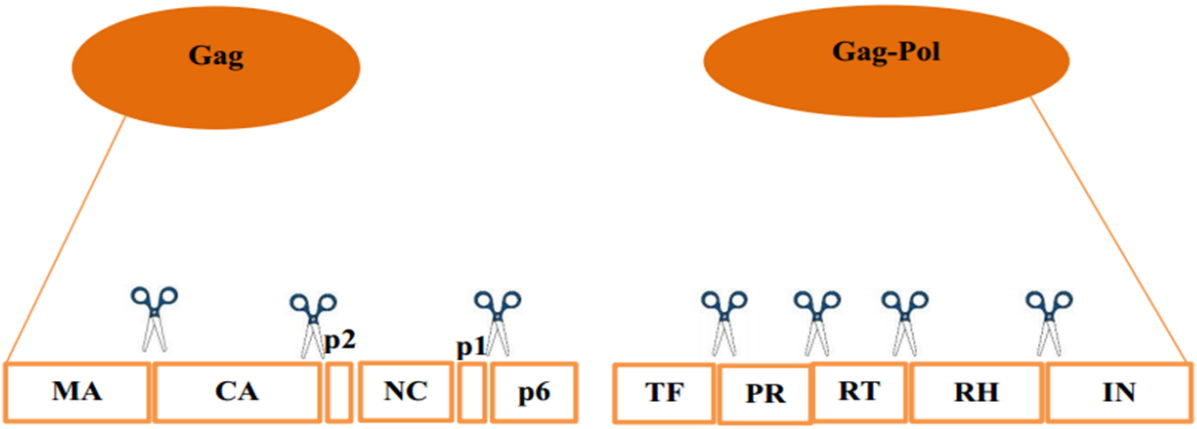


Figure 1: Illustration of Gag and Gag-Pol cleavage process by HIV-1 PR.

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.*[9] in the Protein Data Bank (PDB) with a resolution of 2.0 Å (PDB code 1F7A[10], Figure 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.[9] Those structures are relatively short covering P3-P3′ epitope and most of which are not a close representation of the actual natural substrate segments.[9] Since the inception of this study[9], 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.

Table 1: The nine recognition non-homologous natural substrate polypeptides segments cleaved by the HIV-1 protease.[9, 11]

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
Auto proteolysis site	
Pro-Gln-Ile-Thr-Leu*Trp-Lys-Arg-Pro-Leu	AutoP

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.[9, 11]

Experimental investigations on the interaction between HIV-1 PR and the natural substrate includes: how this homodimeric enzyme recognizes asymmetric substrates,[9, 11] natural substrates mimics for the design of potent HIV-1 PR inhibitors,[12, 13] crystallization of HIV-1 PR—natural substrate complexes,[9, 11] mutations at cleavage sites,[14] and the determination of this enzyme's processing rate of its substrates.[15] 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.[2, 16-19] 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.

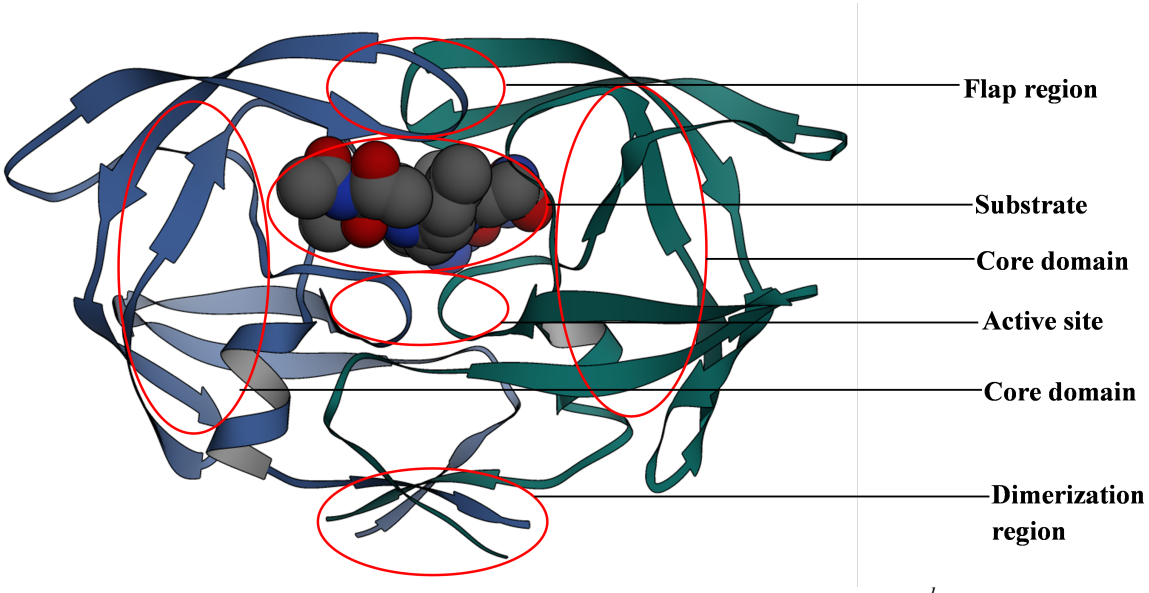


Figure 2: Structure of HIV-1 protease complexed with CA-p2. PDB code 1F7A[1] (RCSB PDB accessed on 28 March 2017). Image created with UCSF Chimera.[3]

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 modelling 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.[3, 20-22] In their 2000 report, Gulnik *et al.*[20] 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[20] 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.

In 2003, Brik and Wong[23] 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:[23] 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[21] 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[24] alongside other aspartyl protease families. This contribution[24] 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 results of titration curves and protonation state at a given pH.[24]

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 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.[25] 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 (Figure 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 recognised and cleaved by HIV-1 PR to produce corresponding

peptide segments of the natural substrates (Table 1).[9, 11] The mechanism of substrate hydrolysis by this PR has remained a major backbone of inhibitor design[26-28] for the treatment of HIV/AIDS.[3, 29]

2.2 Cleavage points

Provided in Figure 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 was mostly associated with substrate modulation, conserved substrate shape, interdependence conformational adaptability of both HIV-1 PR and substrates.[9, 11, 30-35] 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.[36] 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).[37]

2.3 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.[9] The authors noted an extensive binding landscape in which the asymmetry peptide forms 24 hydrogen bonds which does not account for substrate recognition. They however, linked the ability of HIV-1 PR to identify its substrate with 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.[9] An addendum to this, research provides another intriguing investigation where the same authors[9, 11] determined the X-ray crystal structures of five HIV-1 PR—substrate complexes[11] (deposited in the PDB).[10] The sequences of the investigated substrates consists of ten asymmetric amino acid units, which covers 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.[9, 11]

Weber and coworkers[38] 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 modelling 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.[38]

Light emitting holoprotein was also employed to detect the proteolytic bond cleavage site of the MA-CA precursor.[25] 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.[25] 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 HIV-1 protease.[25]

In 2006, Schiffer and coworkers[32] 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.[39] 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.[32] Again, it was resolved that HIV-1 PR identifies the substrate shape in its entirety with little or nil cognizance of its specific sequence.[32]

A closely related research initiative is an all atom simulation study by Perez *et al*.[36] 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)[40-42] 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.[36] The authors[36] estimated the binding 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.[36] (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.[36]

The distinctive binding scope of HIV-1 PR—substrate complex (referred to as "the substrate envelope") was redefined[43] using molecular dynamics (MD) simulations.[44] This study was a sequel investigation from earlier reported work[30, 45-48] 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.[30] Özen *et al*.[43] 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.[43]

Schiffer and coworkers[49] 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[49] 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 remodelled proteins. For side-chain optimization, these authors have used Dunbrack and Cohen-based backbone-dependent rotamer library[50] and the dead-end elimination (DEE) theorem was used for the design of individual pockets.[49]

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,[15, 51, 52] and identification of efficiently cleaved substrates of HIV-1 PR.[53] It was claimed that identification of substrate cleavable sites could serve as the best templates for the synthesis of the better binding inhibitors.[53] 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 cleavage events.[52] Using fluorescein arsenical hairpin (FlAsH) reagent to label a natural substrate containing MA and the N-terminal domain of CA, Swanstrom's group[15] studied the effect of substrate context on processing the HIV-1 PR Gag and Gag-Pol polyproteins. These authors[15] 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.[51]

2.4 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,[54] and 69 HIV-1 PR mutant crystal structures are found in the RCSB PDB.[10]

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,[55] and it was observed that most active site mutations are rarely in contact with substrates, but are crucial to the binding of inhibitors.[30] 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[4]) 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.[30]

The output from this computational simulation[30] was followed by a crystallographic investigation conducted at 1.44–2.10 Å resolution.[31] 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.[31] 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 (Figure 3).[30] Amazingly, their[31] 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.[31] P3-P1 residues of the substrate was also observed to be of great importance in substrate recognition.[30, 31]

Tie *et al*.[56] 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 is 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.

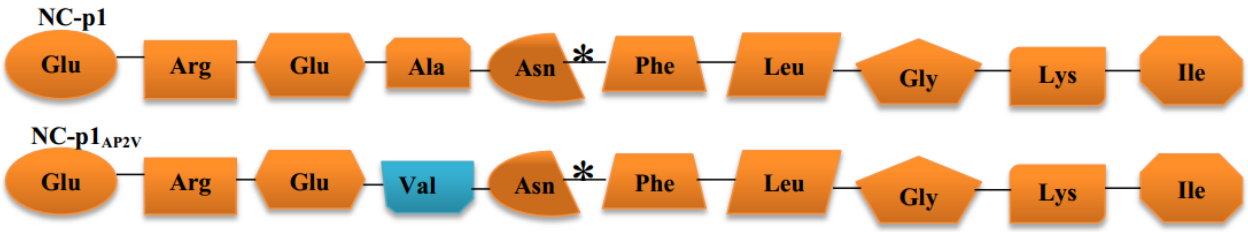


Figure 3: Point mutation at P2 amino acid in NC-p1 natural Gag polyprotein cleavage domain. The scissile bond is indicated with asterisk and mutated amino acid is in green.

In order to confirm this coevolved substrate with mutant HIV-1 PR and to examine in detail their substrate recognition pattern, Schiffer and coworkers[57] carried out another investigation at the molecular level. 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.[58, 59] The molecular modelling and dynamics analysis revealed that the conserved envelope conformation is an important feature for the HIV-1 PR

mutants to recognise and cleave their substrates.[57] Closely related research was again reported by the same group[60] using an Anisotropic Network Model (ANM).[61] In addition to the substrate envelope theory with respect to drug-resistant HIV-1 PRs and coevolved substrates, these authors[60] 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[62] 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.[62] 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[62] a condensed summary is provided by 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:[63] (1) The S-groove can be described as an active pocket within the HIV-1 PR contributed by the two dimers, which allows 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 substrate, a similar bioactivity was observed for both native and MDR PRs.[62]

Lately, the nine cleavage domains of the natural Gag and Gag-Pol precursors were co-crystallized with MDR769 HIV-1 PR by Kovari and coworkers.[63-66] In one of their studies, a wide-open flap and an extended substrate pocket was observed in MDR769 HIV-1 PR.[64] This MDR PR has ten of its residues mutated and each natural target has seven amino units from P3-P4′.[64] It was argued that the enzyme recognizes its substrates through their preserved shape facilitated by conserved water molecules and hydrogen bonds.[64, 66] It was noted that the essential water molecule between the substrate and flap tips (Figure 2) was absent in the reactive region of MDR769 HIV-1 PR.[64] These prompted the authors[66] to postulate the probability of variations in enzymatic mechanism of these mutant with respect to the WT PR.

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.

3 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 literature for HIV-1 PR and its natural targets were documented.[20] 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.

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 ($mM^{-1}{\cdot}s^{-1}$)	K_d
MA-CA	SQNY*PIVQ	WT		3.75[67]	23.00[67]	6.20[67]	
				0.78[68]	15.70[68]	20.10[68]	
MA-CA	VSQNY*PIVQ	WT		0.15[69]	6.90[69]	46.00[69]	
				[38, 68]	6.80[38]	45.30[70]	
MA-CA	QNY*PIVQ	WT	-8.33[64]	0.91[68]	1.60[68]	1.81[68]	8.20e-7[64]
MA-CA	VSQNY*PIV	WT		0.12[68]	7.90[68]	65.80[68]	
MA-CA	SQNY*PIV	WT		0.53[68]	13.50[68]	25.40[68]	
MA-CA	QNY*PIV	WT		0.86[68]	1.00[68]	1.20[68]	
MA-CA	VSQNY*PIVQ	M46L		0.56[69]	18.40[69]	32.90[69]	
MA-CA	VSQNY*PIVQ	V82S		1.34[69]	13.40[69]	10.00[69]	
MA-CA	VSQNY*PIVQ	V82A		0.42[69]	7.00[69]	16.70[69]	
MA-CA	VSQNY*PIVQ	I84V		1.02[69]	93.50[69]	92.00[69]	
MA-CA	VSQNY*PIVQ	L90M		0.64[69]	33.00[69]	51.60[69]	
MA-CA	QNY*PIVQ	MDR769	-2.02[64]				3.40e-2[64]
CA-p2	ARVL*AEAM	WT		0.37[67]	2.60[67]	6.90[67]	
CA-p2	RVL*AEAM	WT	-9.14[64]				2.10e-7[64]
CA-p2	RVL*AEAM	MDR769	-2.02[64]				3.40e-2[64]
p2-NC	ATIM*MQRG	WT		3.00[67]	41.00[67]	14.00[67]	
p2-NC	TIM*MQRG	WT	-9.14[64]				2.10e-7[64]
p2-NC	TIM*MQRG	MDR769	-5.24[64]				1.50e-4[64]
NC-p1	ERQAN*FLGKI	WT		0.17[69]	0.15[69]	0.90[69]	
NC-p1	ERQAN*FLGKI	V82A				0.70[69]	
NC-p1	ERQAN*FLGKI	I84V					
NC-p1	ERQAN*FLGKI	M46L				0.30[69]	
NC-p1	ERQAN*FLGKI	V82S				0.40[69]	
NC-p1	QAN*FLGK	WT	-6.74[64]				1.20e-5[64]

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

NC-p1	QAN*FLGK	MDR769	-3.83[64]				1.60e-3[64]
p1-p6	PGNF*LQSR	WT		>5.00[67]	<0.25[67]	0.05[67]	
p1-p6	RPGNF*LQSRP	WT		1.20[69]	0.98[69]	0.80[69]	
p1-p6	RPGNF*LQSRP	V82A				1.20[69]	
p1-p6	RPGNF*LQSRP	V82S				1.20[69]	
p1-p6	RPGNF*LQSRP	I84V				1.20[69]	
p1-p6	RPGNF*LQSRP	M46L				0.60[69]	
p1-p6	RPGNF*LQSRP	L90M				2.10[69]	
p1-p6	GNF*LQSR	WT	-5.33[64]				1.30e-4[64]
p1-p6	GNF*LQSR	MDR769	-3.52[64]				2.70e-3[64]
TF-PR	FQF*PNIT	WT	-8.84[64]				3.50e-7[64]
TF-PR	FQF*PNIT	MDR769	-3.02[64]				6.30e-3[64]
PR-RT	LQF*PISP	MDR769	-3.42[64]				3.20e-3[64]
RT-RH	AETF*YVDG	WT	-6.91[71]	0.19[67]	1.80[67]	10.00[67]	6.21e-6[71]
RT-RH	ETF*YVDG	WT	-6.44[64]				2.09e-5[64]
RT-RH	ETF*YVDG	MDR769	-2.23[64]				2.40e-2[64]
RH-IN	RKIL*FLDG	WT		1.15[67]	5.10[67]	4.40[67]	
RH-IN	KIL*FLDG	WT	-9.68[64]				6.40e-8[64]
RH-IN	KIL*FLDG	MDR769	-6.53[64]				1.70e-5[64]

Note: PR is the protease, ΔG_{bind} is the intrinsic binding free energy, K_m is the Michaelis constant, K_{cat} is the catalytic constant, K_{cat}/K_m is the rate of catalysis, K_d and K_i denote the dissociation and inhibition constants, respectively.

4 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 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.

4.1 Effect of protonation state on the catalytic aspartate dyad

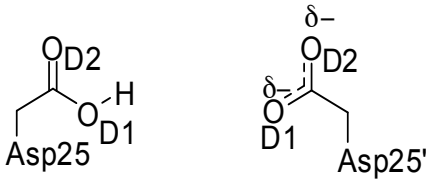
Placement of proton(s) on the Asp dyad of HIV-1 PR is crucial to understanding the reaction mechanism and is still a subject of debate.[2, 72] A review describing the protonation state of the ionisable group of aspartate proteases has appeared recently.[24] Experimental techniques such as, combined X-ray/neutron crystallography,[73, 74] NMR spectroscopy,[75-77], calorimetric and binding

kinetics[24] were used to determine the protonation state in aspartic proteases. Similarly, theoretical methods have also been applied to understand the ionisation state of these catalytic Asps. These methods include; QM/MM,[18, 78-80] MD simulations[81-83] and online software (such as PROPKA,[84] PDB2PQR[85] and H++[86]). One of the latest theoretical reports dedicated to detailed protonation pattern of the Asp dyad is from Roitberg and coworkers,[87] where pH Replica-exchange MD (REMD)[83] 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 ionisable 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.[87] A correct understanding of the protonation state(s) of the HIV-1 PR—substrate complex catalytic aspartates is crucial for accurate QM modelling of reaction mechanism.

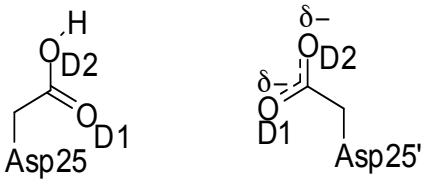
4.2 Possible conformation of HIV-1 PR for catalytic function

Evidence from both experimental and theoretical analyses has revealed that the catalytic process of HIV-1 PR functions at pH of 2–7.[23, 24, 88] This process usually occurs at the active site, which consist of two coplanar aspartate moieties and is usually stabilized by the catalytic triad Asp25(25′), Thr26(26′) and Gly27(27′) peptide units.[89] The Asp25(25′) dyad of the free enzyme can exist in monoprotonated[87] (four possibilities, model **A-D**), an unprotonated[90, 91] (model **E**) and diprotonated[92, 93] (four possibilities, model **F-I**) forms (Figure 4).

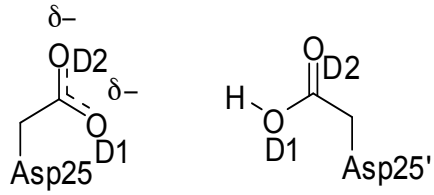
Piana and Carloni[88] reported the observation of a low barrier hydrogen bond (LBHB) within the apo HIV-1 PR of reactive Asp monomers (model **J**, Figure 4). These researchers[88] utilised *ab initio* simulations at 300 K to highlight the requirements for this enzyme to carry out its catalytic functions 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 stability of the Asp dyad orientation.[88] Furthermore, the catalytic reactivity of HIV-1 PR—substrate complex is dependent on the conformational flexibility of the enzyme.[17, 94, 95] 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 is recommended to obtain reasonable geometry orientation and energetics.[1, 17]



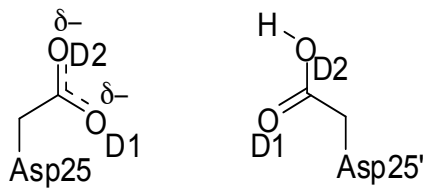
Model A; OD1 of Asp25 is protonated



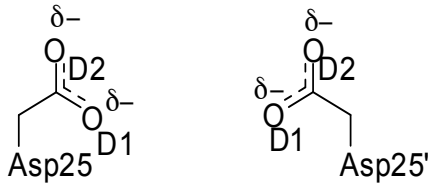
Model B; OD2 of Asp25 is protonated



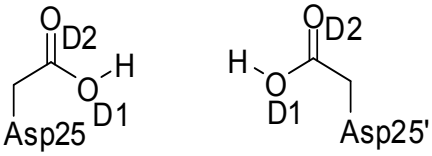
Model C; OD1 of Asp25' is protonated



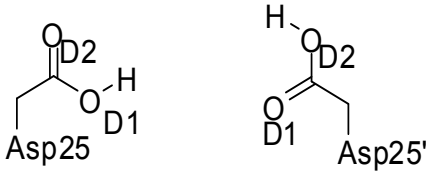
Model D; OD2 of Asp25' is protonated



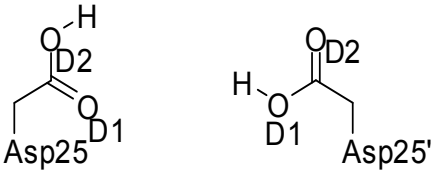
Model E; Unprotonated



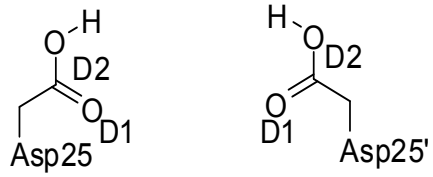
Model F; OD1 of Asp25 and Asp25' are protonated



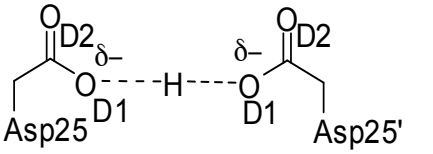
Model G; OD1 of Asp25 and OD2 of Asp25' are protonated



Model H; OD2 of Asp25 and OD1 of Asp25' are protonated



Model I; OD2 of Asp25 and Asp25' are protonated



Model J; Low barrier hydrogen stabilizes OD1 of Asp25 and Asp25'

Figure 4: Proposed protonation models for the HIV-1 PR catalytic dyad.[4-7]

The catalytic core domain of the PR must attain a specific geometry before it can process its substrate/inhibitor's functionality, Okimoto *et al.*[81] 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 (Figure 5). Using the Hartree-Fock (HF)[96] theoretical method and the 6-

31G(2d)[97] basis set, these authors[81] determined the necessary conditions for an active conformation to enable HIV-1 PR cleave its natural substrates. These conditions are: (a) the two catalytic Asp dyad should hold WatC through formation of hydrogen bonds (Figure 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 (Figure 5: d5) and this was based on the VDW radii of the carbon and the oxygen in retrospect,[81] (c) and the oxygen of the cleavable carbonyl forms a hydrogen bond with the OD2 of Asp25′ (Figure 5: d6).

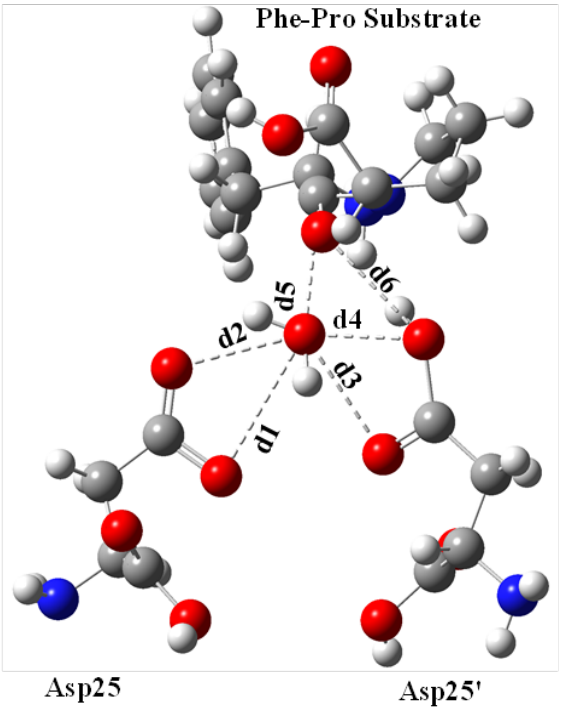


Figure 5: Interatomic distances obtained with *ab initio* HF/6-31G(2d) for the ES complex, redrawn from literature.[8] Carbon, oxygen and nitrogen atoms are represented with grey, red and blue, respectively. All distances are in Å, d1=3.28, d2=2.66, d3=2.96, d4=2.80, d5=2.91 and d6=2.97.

The effect of conserved water molecules on the reactive conformational state of the enzyme was also addressed therein.[14, 98] The 3-D arrangement of dipoles from these structurally conserved water molecules contribute to the favourable geometric of this enzyme required for its cleavage bioactivity.[98]

4.3 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, 99, 100] or a general acid-base[1, 99, 101-108] reaction. The major distinction between these two mechanisms is the presence of 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.

4.3.1 Nucleophilic reaction

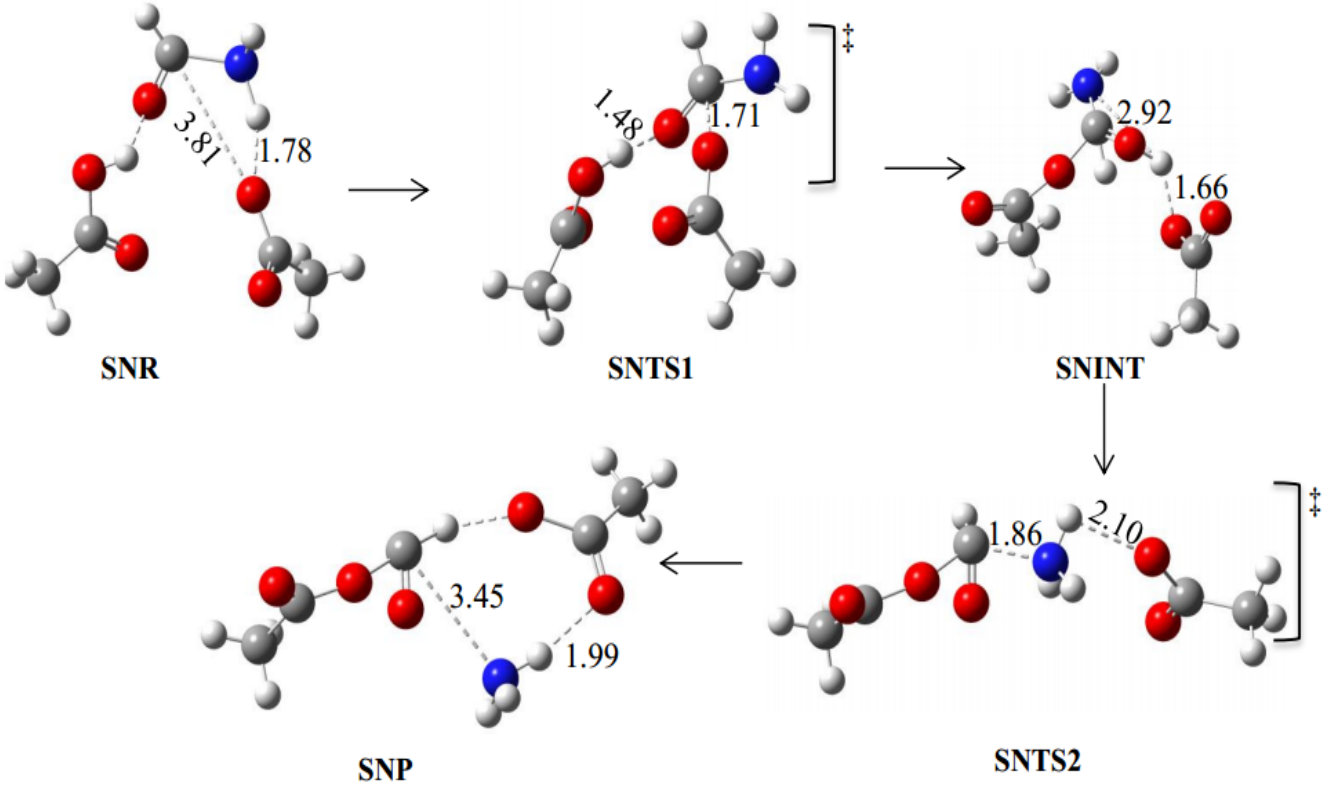
Earlier thoughts about the hydrolysis of natural substrates by HIV-1 PR was that it most likely proceeds through direct nucleophilic attack by an aspartate.[99] This direct nucleophilic attack process involves the catalytic aspartate group whereby the unprotonated Asp acts as the nucleophile rather than 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.

4.3.1.1 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 hydrolysed, with the unprotonated Asp acting as general base.[99] The stepwise nucleophilic mechanistic pathway was explored herein[1, 99, 100] 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 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)[109] method was parameterized and tested on the cleavage of the natural substrate by the HIV-1 PR.[110] The AVB formalism was developed based on density functional theory (DFT)[111] 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 centre[110] were modelled using AVB while the rest of the system was treated with a classical MD[44] method. This computational method allows the authors[110] to investigate the stepwise nucleophilic

mechanism in which protonation was assigned to the two Asp group (model **I**, Figure 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,[110] it was impossible to estimate the energy barriers of the entire nucleophilic process.

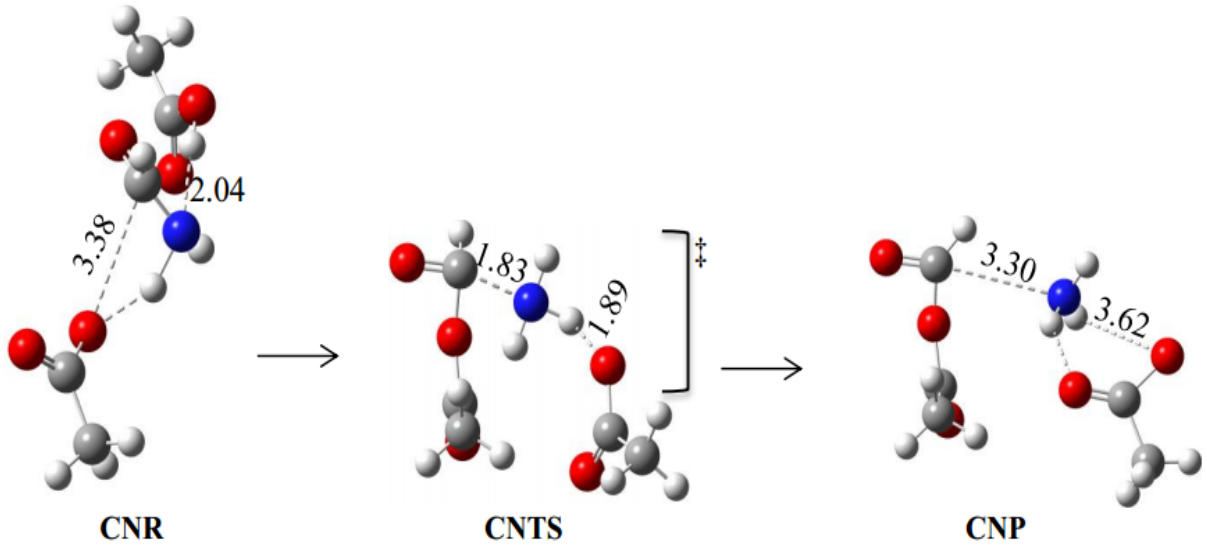


Scheme 1: A pictorial representation of the structures involved in stepwise nucleophilic mechanism for the hydrolysis of a simplified peptide model by HIV-1 PR using MP2/6-31G(2d)//RHF/6-31G(2d).[9] The legends are: SNR = stepwise nucleophilic reactant, SNTS1 = stepwise nucleophilic transition state 1, SNINT = stepwise nucleophilic intermediate, SNTS2 = stepwise nucleophilic transition state 2, SNP = stepwise nucleophilic product. All distances are in Å.

4.3.1.2 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 Asp) on the carbonyl carbon and transfer of 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-order perturbation (MP2)[112] and HF[96] theories coupled with the 6-31G(2d)[97] 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]



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).[9] The legends are: CNR = concerted nucleophilic reactant, CNTS = concerted nucleophilic transition state, CNP = concerted nucleophilic product. All distances are in Å.

4.3.2 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[113, 114] following previous experimental research on related proteolytic enzymes.[115, 116] Recent evidence weighs strongly in favour of the general acid-base catalysis through the use of combined X-ray/neutron crystallography,[73, 74] 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, 117] 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.

4.3.2.1 Stepwise general acid-base mechanism

The stepwise general acid-base 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

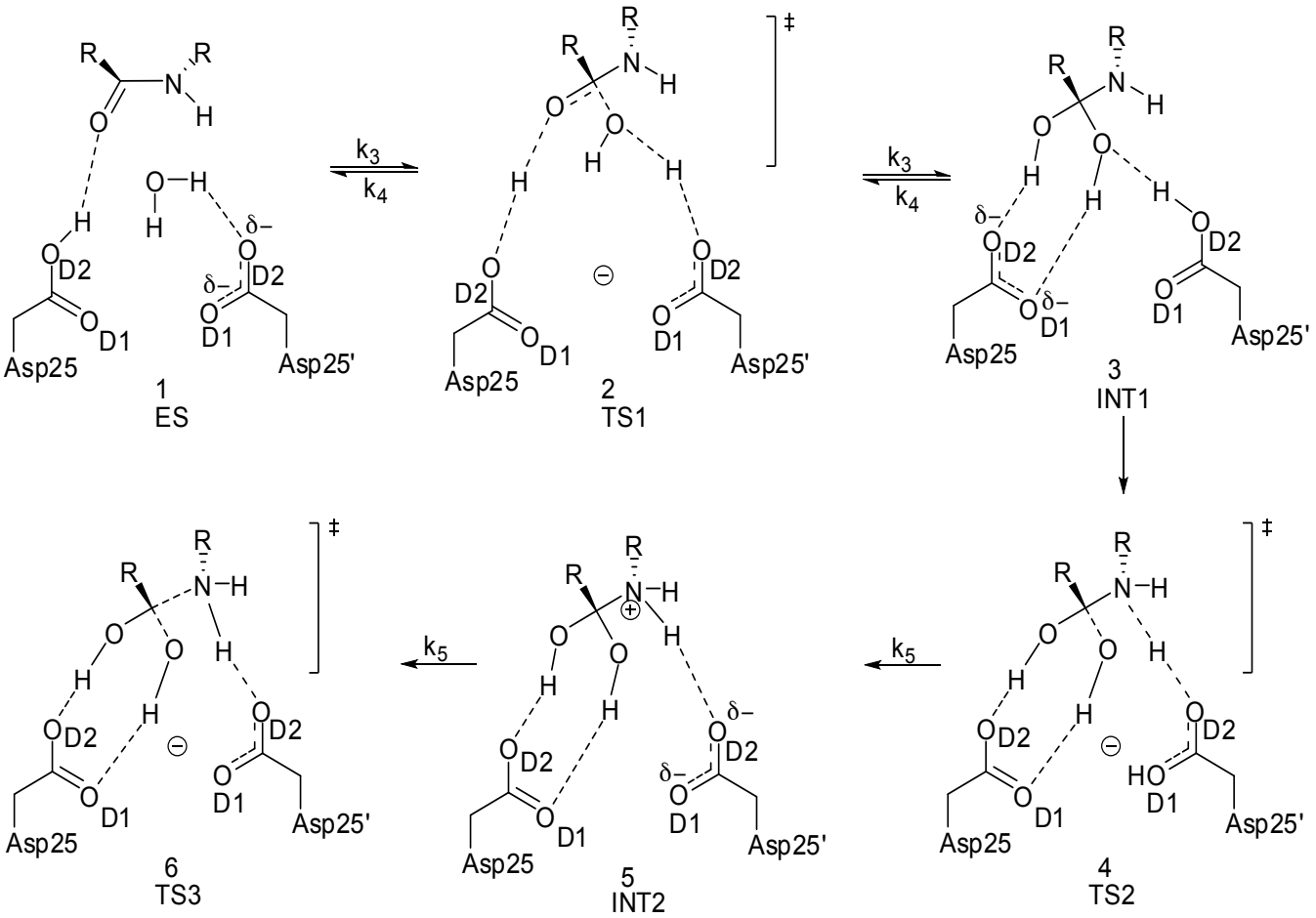
questions are: Can a single protonation pattern be sufficient for the proteolysis of all the natural substrates? How many TS structures are involved in stepwise general acid-base mechanism and what is the rate-determining step?

Within the framework of this review, we have pull studies together based on the protonation pattern (Figure 4) used/proposed by the different contributors from both computational and experimental investigations. This becomes a better approach due 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**; Figure 4).

4.3.2.1.1 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 (Figure 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)[109] method was parameterized and tested on the cleavage of the natural substrate by the HIV-1 PR.[110] Protonation of the catalytic groups was specifically on OD2 of Asp25 (model **B**, Figure 4). Shown in Scheme 3 is the adopted reaction model by Trylska *et al*.[110] 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 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.[110] 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.[110]



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.[10-12]

Trylska *et al.*[100, 107, 110] 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** (Figure **4** and Scheme 3) was found to be the most suitable model for the general acid-base enzymatic catalysis of HIV-1 PR on *N*-acetyl-Thr-Ile-Met*Met-Gln-Arg-amide sequence.[100] 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;[100] 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 was 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[100] suggested a technique that explores the full conformational space such as umbrella sampling.[118, 119] 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,[94, 95] Rothlisberger and coworkers[17] offered a hybrid Car-Parrinello (quantum mechanics) and molecular mechanics (QM/MM) method[120] 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[121, 122] DFT formalism and the plane wave (PW)[123] basis set for the QM part, these authors[17] investigated two possible protonation patterns (models **B** and **J**, Figure 4) for the ionisable 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 (Figure S**1**; SI), which was proposed to be possible through a tunnelling[124] 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.[17] The $\Delta G^{\ddagger}$ for the formation of gem-diol INTs was 18 kcal mol^{-1} *via* model **B** protonation mechanism. The data suggests that the catalytic reaction is more favourable when OD2 of Asp25 is protonated. Even when the reverse reaction was examined the same free energy barrier was obtained,[17] and including the catalytic triad in the QM part to monitor the reaction mechanism gave activation free energy of 15 kcal mol^{-1}.[17] 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.[95] 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},[17] an indication of the vitality of flexibility for reactive catalysis.[17, 95] 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.[125, 126] These authors[17] also carried out QM calculations to rationalize this discrepancy.

Two of our hypothetical questions form part of Bjelic and Åqvist[98] aims when they analysed 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)[127, 128] was coupled with free energy perturbation (FEP)[127, 129] 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.[98] After detailed literature survey, protonation pattern **B** (Figure 4) and the general acid-base stepwise mechanism (Scheme 3) were adopted for their[98] 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,[98] in which the TS1 value is in agreement with experimental value of 17.6 kcal mol^{-1}.[101] Apart from the active site WatC, a total of five water molecules was observed to be crucial for stability and reactivity,[98] which is supported by an earlier crystallographic study of this enzyme.[14] 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}.[98] 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.[17, 130]

Kipp *et al*.[131] gave an answer to one of research questions on the similarity between the mechanistic route of drug-resistant and the native HIV-1 PRs catalysis of substrates. They[131] 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.[132] Kinetic Isotope Effects (KIEs)[114] 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[131] to have occurred in six phases (Scheme 3) and TS2 was noted as the rate-determining structure as earlier proposed in literature.[17, 101, 106] This TS corresponds to the protonation process of the substrate scissile nitrogen atom.

The relevance and implication of computational techniques in studying reaction mechanism

cannot 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 advanced computational technique such as enhanced sampling that explores the full conformational space of the reaction process.[100] Krzemińska *et al*.[108] recently investigated 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 (Figure S1; SI). Although the modelled system is a 6-Alanine peptide sequence (not an exact natural substrate), some highlights from their[108] 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**; Figure 4) leading to 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[101] with $\Delta G^{\ddagger}$ of 15 kcal mol^{-1}.[108] This theoretical $\Delta G^{\ddagger}$ is in agreement with experimental values within 15 and 17.9 kcal mol^{-1}.[113, 133] A concerted process was observed for the plausible rate-determining step, which is characterized with two proton transfers. The first involves 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[108] also measured the enzyme 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[134]) 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.[108]

4.3.2.1.2 The stepwise general acid-base mechanistic pathway of HIV-1 PR involving Asp25′ protonation

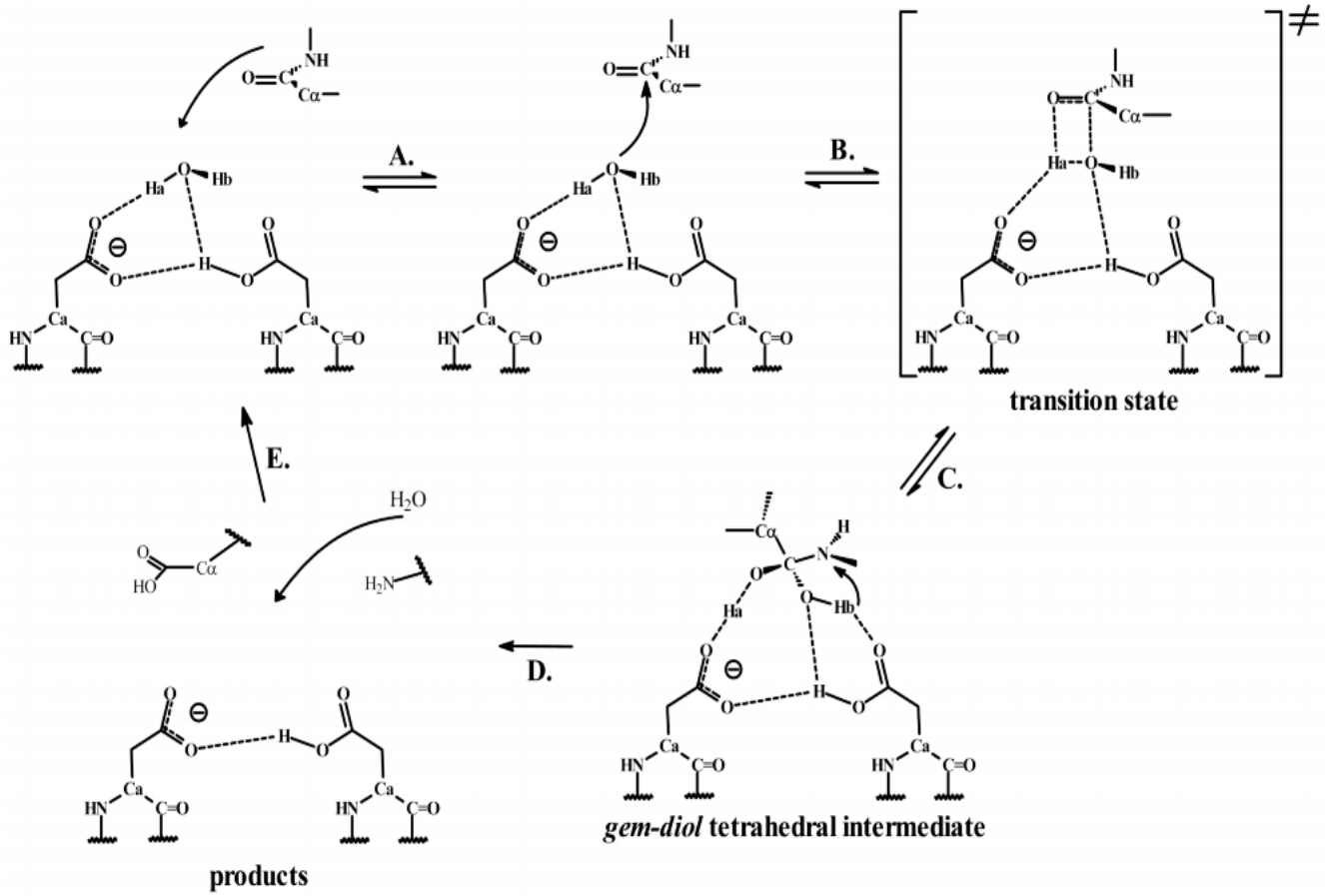
The catalytic ability of HIV-1 PR was traced to its conformational adaptability by Piana *et al*.[95] They[95] employed an integrated approach with *ab initio* Car-Parrinello MD simulation[120] (10.0 ps) and classical MD simulations (2.1 and 8.0 ns) at 300 K to investigate the HIV-1 PR—N-acetyl-Thr-Ile-Met*Met-Gln-Arg complex in which the substrate segments corresponds to six amino acid

residues of p2-NC (Table 1).[95] 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 modelling leading to the tetrahedral intermediate formation.[95] The authors[95] observed a mechanistic pathway which showcase OD1 of Asp25′ protonation (model **C**; Figure 4) as the energetically favourable compared to the rest of the models (**A**, **B** and **D**; Figure 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[95] was 5.5 kcal mol^{-1} higher than experimentally estimated values.[113, 114, 135] These authors[95] however, attributed this large difference in the theoretically predicted energetics to their modelled 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.[122, 136] It was concluded that the catalytic triad of the cleavage sites mainly induced the polarization of the reactants. The classical MD calculation revealed a large-scale protein fluctuation inherent from flaps and the cantilever, which modulates the conformational properties 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.[95]

Weber and coworkers[137, 138] 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.[137, 138] 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 on the nature of the amino-gem-diol INT.[137] Protonation on the Asp dyad was observed to be favourable on OD1 of Asp25′ (model **C**, Figure 4). The PRs neither forms covalent bonds nor transfer its carboxyl hydrogen to the substrates[137] in contrast to Schemes 1-3 as revealed through experiment,[137] which will be discussed subsequently.

The proposed mechanism proceed with the Asp group participating as anchoring moieties (Scheme 4).[137] The reaction process involves formation of an enzyme—substrate complex (step

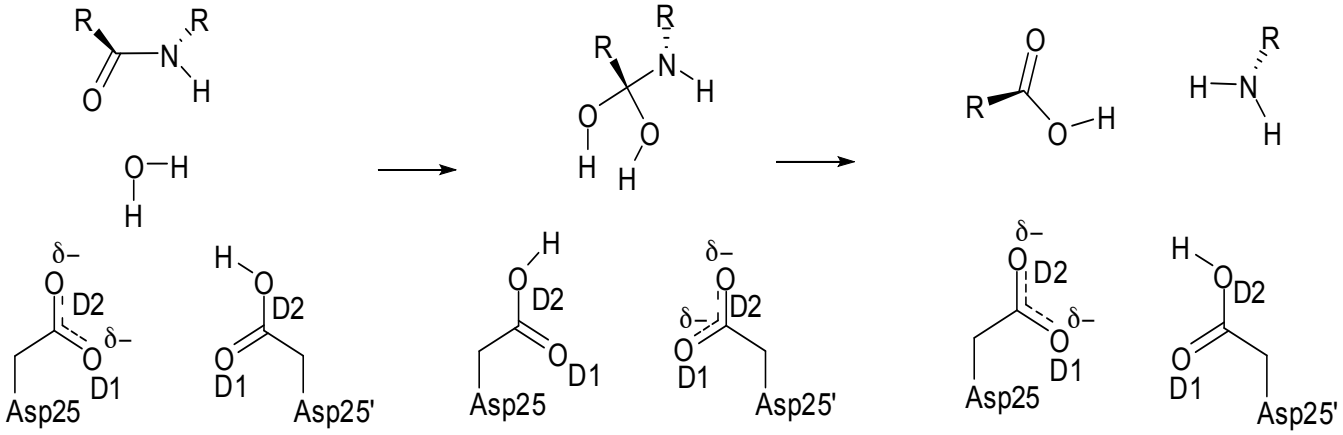
A), nucleophilic attack of WatC on the scissile carbonyl with formation of 4-membered ring TS (step B), formation of gem-diol tetrahedral INT (step C), decomposition of the INT into product complex/protonation of scissile nitrogen (step D) and separation of the cleaved substrate from the enzyme (step E). The decomposition of the INT, which features protonation of the scissile nitrogen, was suggested to be the rate-determining step[137] as proposed from previous theoretical research.[106]



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[13] Copyright (2007) American Chemical Society.

Okimoto *et al*.[81] investigated the mechanism of this enzyme with the natural substrate using MD[44] 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**; Figure 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[81] 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[81] 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.



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.[8]

Next, the same group[81, 82] 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.[82] This was studied using MD simulations in two different approaches; the no cut-off and the Particle Mesh Ewald (PME)[139] methods in which analysis were drawn at 300 K for 100 ps and 300 K for 250 ps,[82] 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.[82] 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 proceed to intermediate formation, substrate protonation and cleavage stages (Scheme 5).[81, 82] 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.[82]

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.*[18] 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.[18] 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[96, 122] with DZVP,[140] 3-21G[141, 142] and STO-3G[143, 144] 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 (Figure 6)[18] which were assigned to AMBER force field (*ff03*).[145] The protonated ionisable group of the reactive aspartates was noted to be Asp25′ OD2 (model **D**; Figure 4 and Scheme 5).

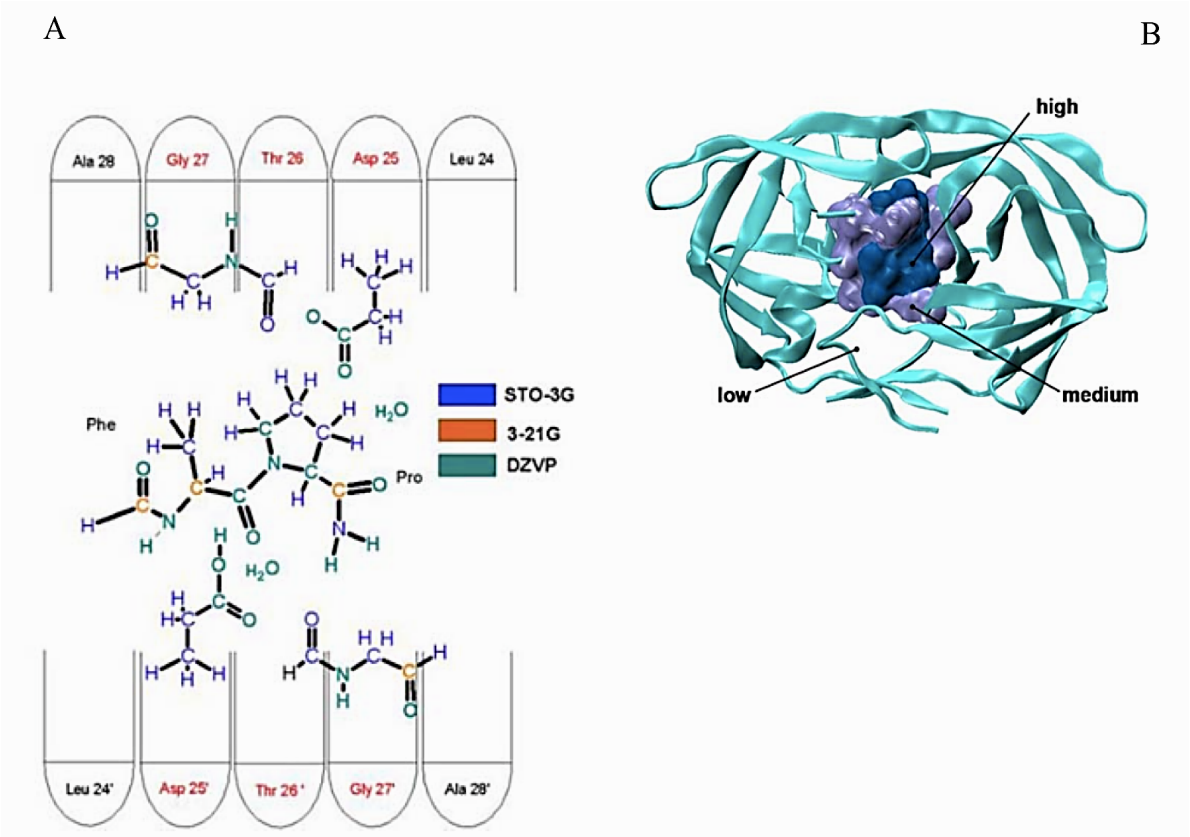


Figure 6: A; The basis set adopted for the atoms of the QM system are shown using different colors. B: Model system used to study the QM/MM reaction profile of the enzyme HIV-1 PR. Pictures are adapted from literature[12] with permission "Copyright (2007) Springer-Verlag".

From the QM calculation, four minima structures and two TS structures were observed along the PES for substrate hydrolysis.[18] One of the studied TS structures (corresponding to TS2 in

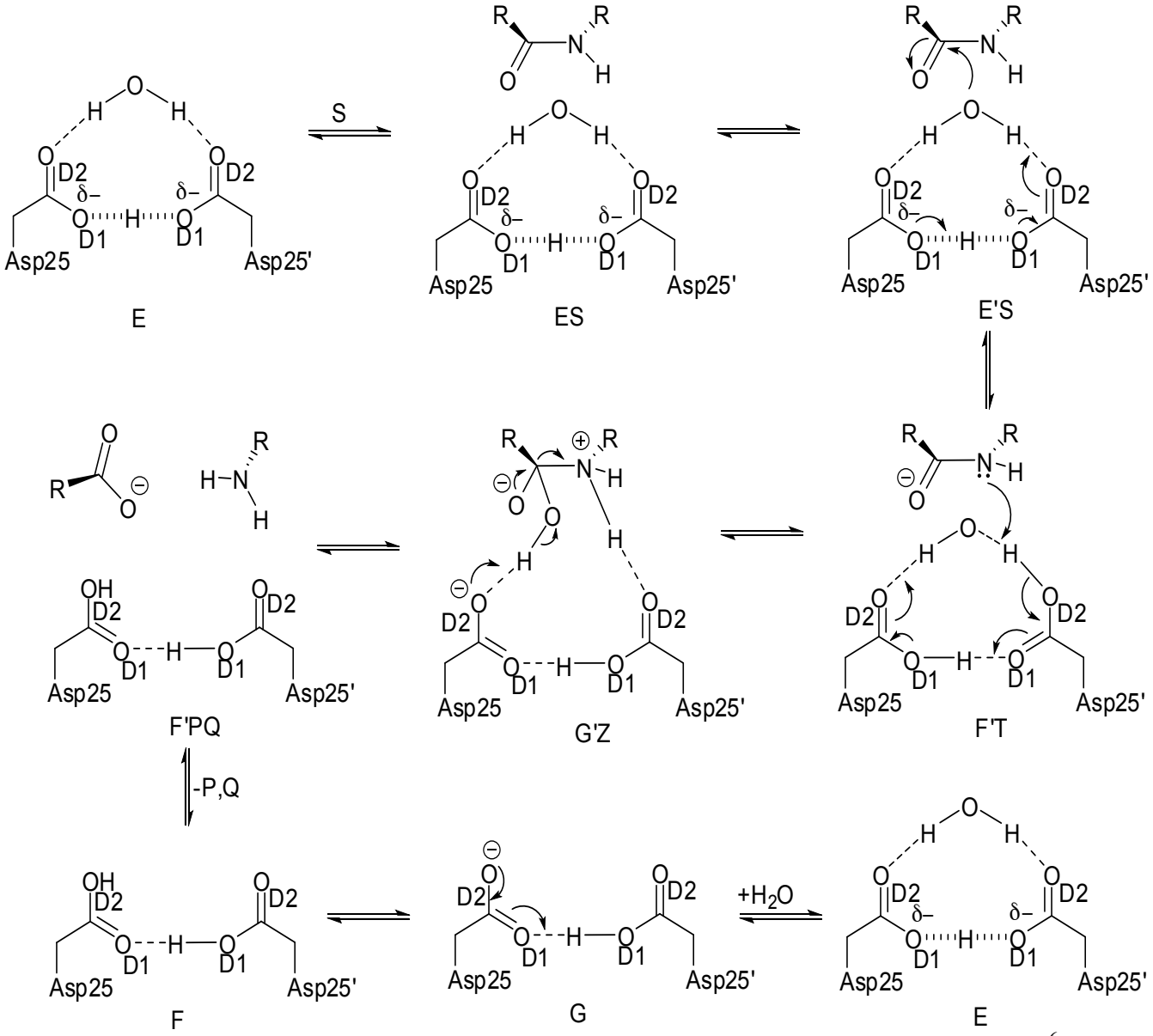
Scheme 3) corresponds to an earlier experimentally[101] and theoretically[106] proposed rate-limiting step with a characteristic highest peak along the reaction coordinate.[18] The activation free energy of approximately 23 kcal mol^{-1} is comparable to experimentally deduced results.[113, 114] The 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[18] 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.[18]

4.3.2.1.3 The LBHB stepwise general acid-base mechanistic pathway of HIV-1 PR

Northrop[124] 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**, Figure 4) in which seven proton transfers were proposed for the catalytic turnover using a chromophoric substrate (Scheme 6). Highlighting other related observations[146, 147] to provide exclusive understanding of the purpose of the LBHB in Asp proteases, Northrop[124] hypothesized that LBHB facilitates the formation of a cyclic complex. Formation of cyclic moieties along the reaction coordinates in a given chemical process has been a subject of theoretical research in our group for over a decade.[79, 148-154] In aspartate proteases, cyclic complexation allows the distribution of electron density which stimulates the rate acceleration through reactant-state tunneling.[124]

In addition to studying other stepwise general acid-base mechanism, Rothlisberger and coworkers[17] investigated the LBHB model. They studied the effect of protonation pattern on the formation of ES complex in which protonation model **J** (Figure 4) was energetically more favourable than model **B** with a difference of 0.5 kcal mol^{-1}.[17] 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.[113, 133] The mechanism takes the form of Scheme 4[137] 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).



Scheme 6: Chemical and kinetic isomechanism of an aspartic protease redrawn from literature.[6] 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.

4.3.2.1.4 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,

[117] An exact natural substrate is cleaved by the active HIV-1 PR before the crystals grow and data collection becomes difficult.[117, 155] However, efforts were made by Hosur and coworkers[117] 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 (Figure 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.[117]

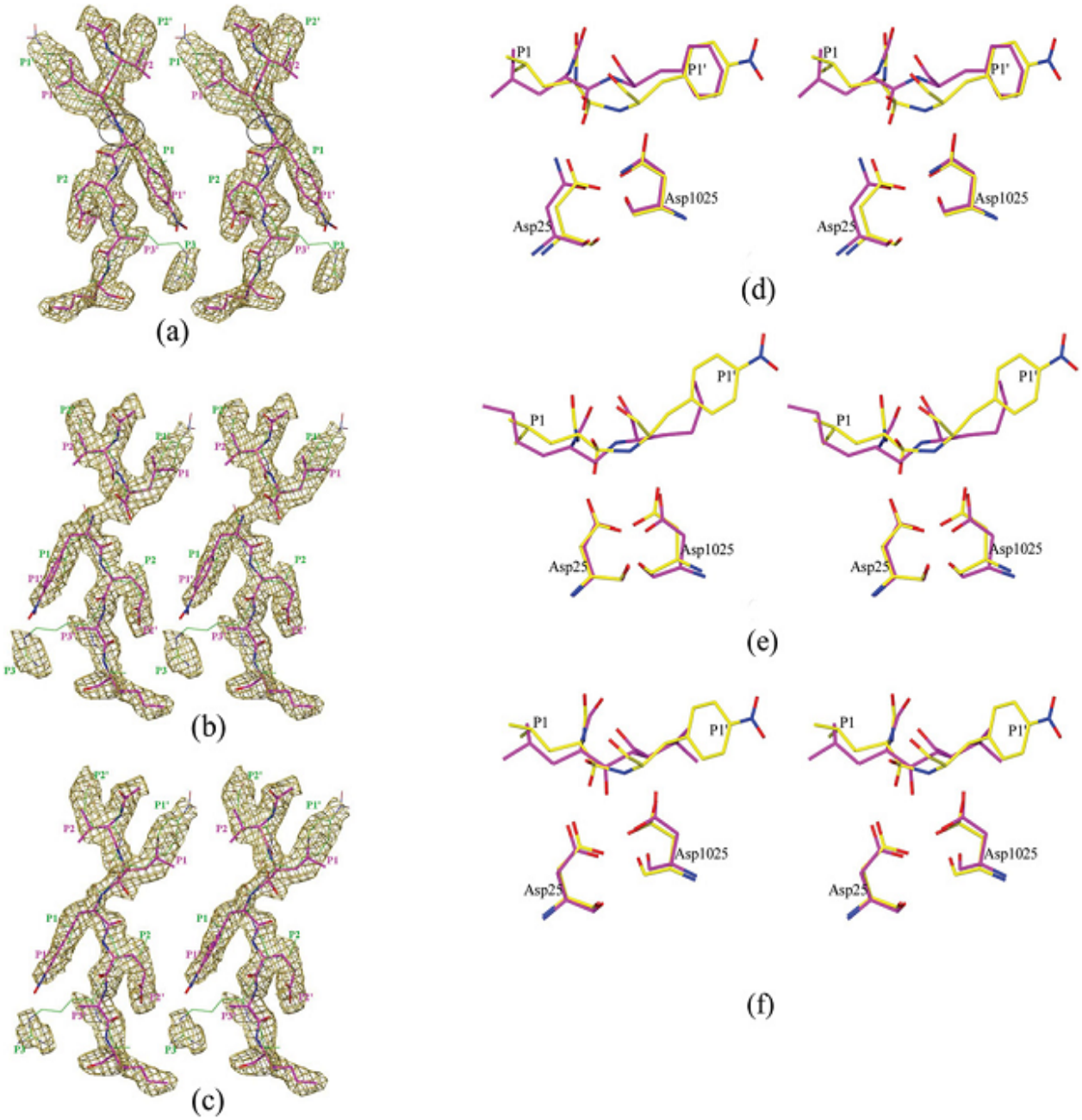


Figure 7: The tetrahedral peptidolysis intermediate. (a)–(c) Stereo diagrams of $2F_o–F_c$ maps (contoured at 0.8σ). (a) The substrate is refined as a regular peptide. The region of the model between the P1 and P1′ residues, which does not fit properly in the density, is circled. (b) The substrate is 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.[14] Image and details are taken from literature[14] with permission "Copyright (2005) Biochemical Society Portland Press".

These authors[117] proposed that the formation of INT and protonation of the scissile peptide nitrogen occur sequentially. A better insight on the nature and conformation of this INT candidate may potentially be useful in the prediction or synthesis of tighter HIV-1 PR inhibitor binders.[117] A similar observation was made from another crystallographic analysis.[156] 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.[157] 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[124] isomechanistic hypothesis (Scheme 6), Das *et al*.[156] noted four hydrogen-bonding interactions between the dyad and the decapeptide substrate. From their[156] observations and discussions on the protonation pattern of the ionisable aspartates through the analysis of the hydrogen bonds at the catalytic centre (Figure 8), the reaction proceeds with protonation model **J** (Figure 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[117, 156] serve as potential answer to the experimental detection of reaction complexes along HIV-1 PR—substrate PES.

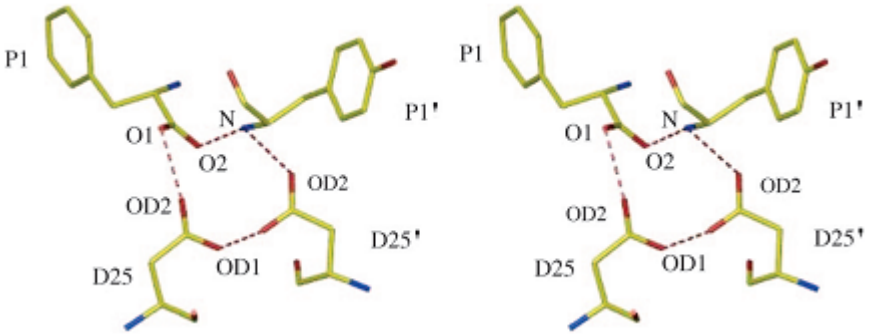


Figure 8: Hydrogen-bonding interactions at the catalytic centre are shown by dotted lines. Picture adapted from literature[15] with permission "Copyright (2006) National Academy of Sciences, U.S.A"

Another attempt to capture reacting structures along the PES of HIV-1 PR—substrate complex was done by Bihani *et al*.[158] The product segments of the hydrolytic action of HIV-1 PR on the matrix and capsid (MA-CA) junction (Table 1) was elucidated with X-ray crystallography refined to 2.0 Å resolution. Using a similar approach described in literature,[117, 156] they[158] have enabled us to summarise the hydrogen bond distances within the reactive Asp dyad and the natural substrate (Table S**1**; SI). Two more important observations were made with respect to hydrogen bonding.[158] 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,[156] 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.[158]

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.[117, 156, 158-161] Using a solution of an undecapeptide substrate at pH 7, Prashar *et al*.[160] 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**, Figure 4) in which the proton is shared between the OD1 of both Asp molecules. Figure 9 shows the proposed mechanistic route for HIV-1 PR and substrate catalysis[160] based on 3-D crystal structures from literature.

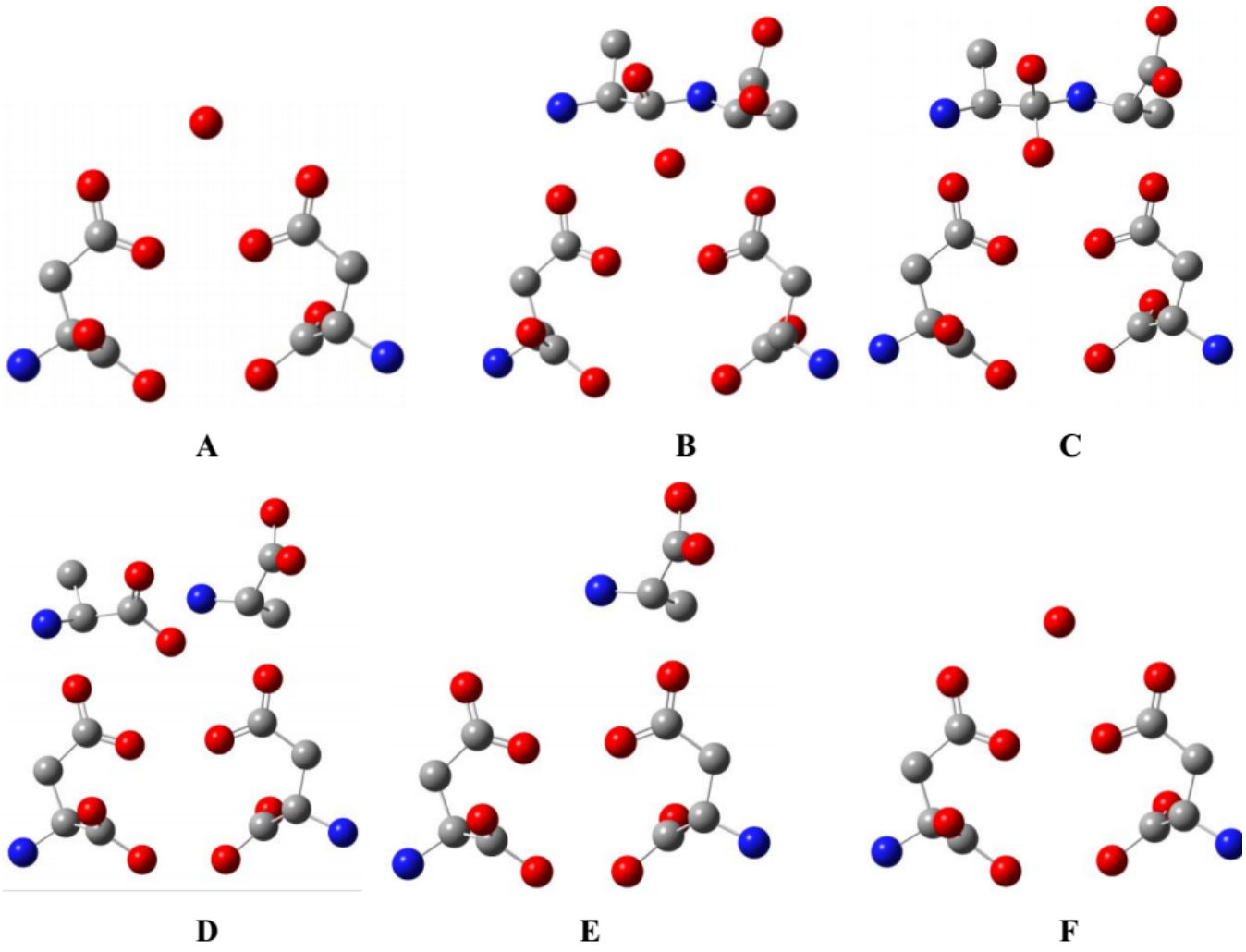


Figure 9: Proposed sequence of steps (**A–F**) in the cleavage of natural substrates by HIV-1 protease redrawn from literature[16] (hydrogen atoms were omitted for clarity). Main chain atoms of substrate peptide and Asp molecules are shown in grey=carbon, blue=nitrogen and red=oxygen). Each figure is based on the structure indicated: (**A**) Bound WatC to ligand-free tethered HIV-1 PR Asp dyad, PDB code 1LV1.[17] (**B**) Modelled ES complex. (**C**) Formation of INT, PDB code 3MIM.[18] (**D**) Breaking of the scissile bond with the product peptides bound in the active site, PDB code 2NPH.[15] (**E**) Diffusion of a unit of the separated products out of the active site and WatC into the active site, PDB code 2WHH.[16] (**F**) Release of the other product segment and movement of WatC into its original position, PDB code 1LV1.[17]

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.[161] 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.[161] Detailed literature review[24, 92, 93] and analysis with respect to relevant work,[78, 162, 163] was also presented.[161] This was followed by a thorough examination of the nature of the hydrogen bond and geometry of the

coplanar Asps. In short, protonation pattern **B** (Figure 4) was upheld to be suitable for the tetrahedral INT description.[161] The proposed mechanism leading to product formation is summarized in Scheme 3. One of the distinctions of this investigation[161] 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 S**1**, 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 is the rate-determining step as observed in literature.[17, 101, 106, 161]

An autoproteolytic peptide sequence cleavage mechanism by WT, I54V, V32I and I47V HIV-1 PR mutants were investigated.[137, 138] To capture notable moieties along the reaction pathway of these drug-resistant—peptide systems, Weber and coworkers[137, 138] 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′.[137] 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 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).[137]

"Snapshots" of three successive steps along the reaction coordinate of V32I and I47V—nonapeptide sequence was obtained through a careful X-ray crystallographic analysis.[138] 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,[138] a related hypothesis from Schramm and coworkers.[131] A reaction mechanism was proposed with a proton placed on OD1 of Asp25′ (to initiate reactivity) in a five-step process (similar to Figure 9) for these enzymes.[138] 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.[138] (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[138] as suggested in a number of studies.[17, 101, 106, 131, 161] (3) The 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.[131] (4) The interactions observed in these new structures[138] 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 a LBHB, of the hydroxyl group of the peptide intermediate with one of the catalytic aspartates, as reported in lower resolution crystal structures.[117, 161]

These contributions[117, 137, 138, 156, 158, 161, 164] 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 enzymatic process through crystallographic analyses is appreciably informative especially when different intermediate structures are discovered.[164] A clearer picture on 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 comprehensive picture of HIV-1 PR reaction.[164]

More recently, neutron crystallography was used to capture proton shift in HIV-1 PR—darunavir system at the reactive site of the enzyme.[74] Variations of Asp protonation states that were earlier proposed[24] were observed for the first time. The catalytic Asp dyad rarely maintain single protonation model throughout the reaction process.[2, 81, 95, 100, 161, 165] Gerlits *et al*.[74] applied neutron crystallography to detect two proton transfers at 2.0 and 2.3 Å resolutions within acidic pH of 6.0 and 4.3, respectively. A shared proton between the OD1 of both Asps (a LBHB) in the substrate-free system moved to OD2 of Asp25 while, proton from OD1 of Asp25′ was transferred to the OH of darunavir.[74]

4.3.2.1.5 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[117, 137, 138, 161, 164] (INT1 Figure 10) compared to the oxyanion INT (INT2 Figure 10), which was proposed to be possible

through a tunnelling effect.[124] INT1 is a typical TS analogue whose conformational mimics has served as basis for the design of nearly a dozen FDA-approved drugs for HIV-1 PR inhibition.[132, 166] In addition to theoretical investigations[17, 107, 108, 165] 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*.[167] 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[167] remarked that the neutral gem-diol INT is highly stable compared to the ionized oxyanion INT (Figure 10); a similar observation noted here.[17, 107, 108] The applied model was the same as the one studied before[17] with the inclusion of the enzyme frame, counterion and explicit solvent to account for the entire system's behaviour in the present model.[167] Illustrated in Figure 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 entropic contribution is considered, the likelihood of change in the instability of this later conformation is uncertain.[167]

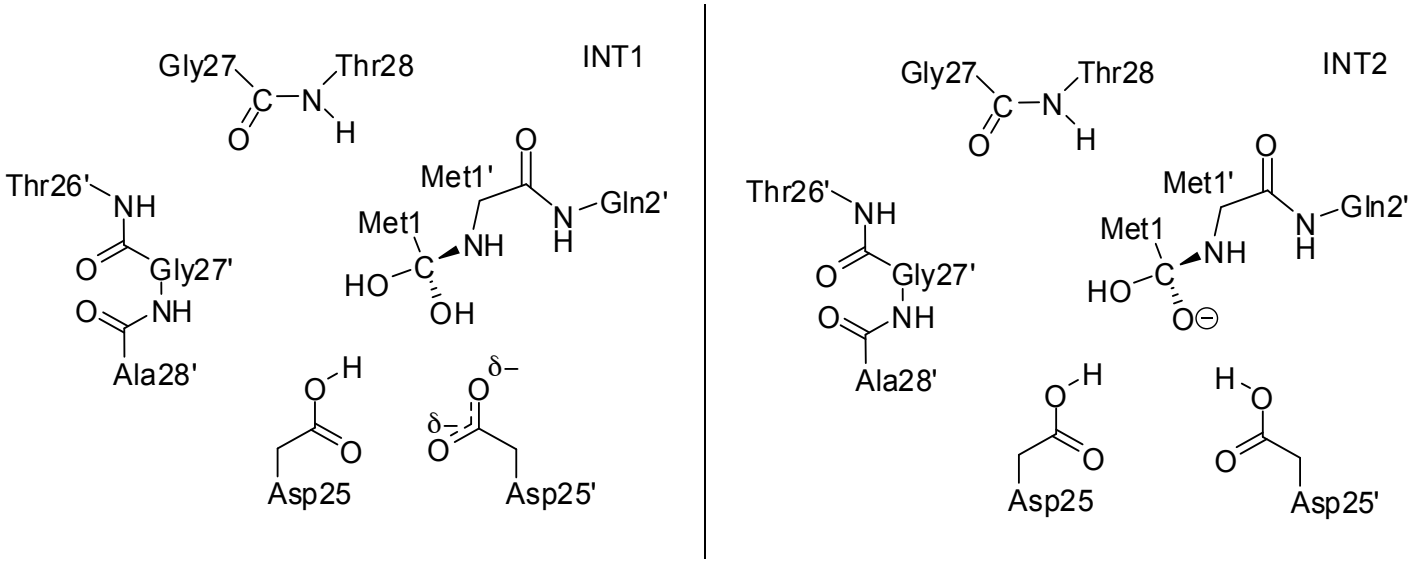
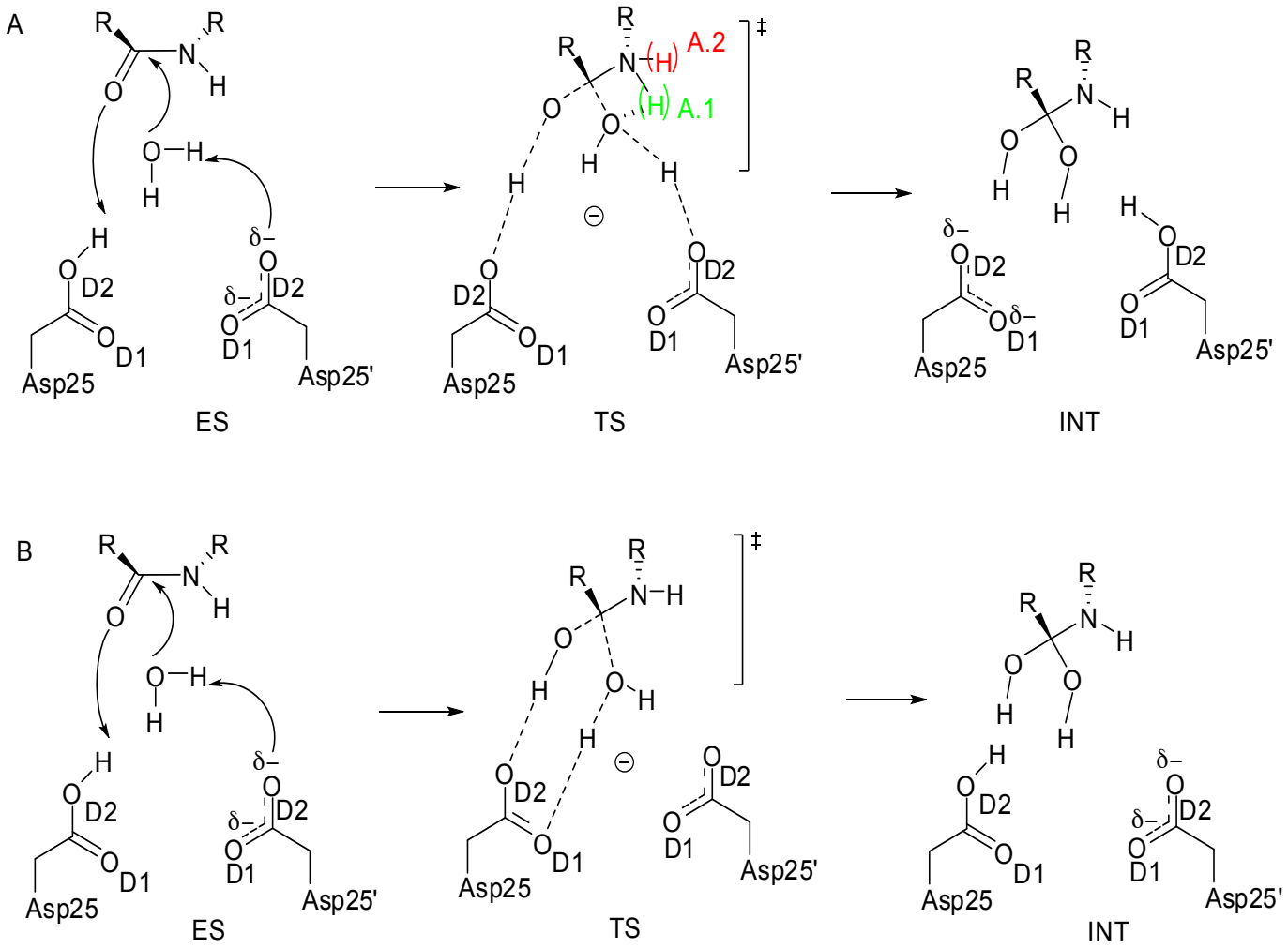


Figure 10: Schematic representation of the INTs; INT1 is the gem-diol intermediate[19] while INT2 is the oxyanion form[20] of the substrate. Both are gas phase models in which the entire systems were partitioned into three layers. The only structural difference between the neutral and the charged reaction 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.[21]

Garrec *et al*.[2] 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** (Figure 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.[2] 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--$^{Asp25\prime}$OD1--$^{Asp25\prime}$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)[168, 169] method, they[2] 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[2] involving many frozen atoms close to the active site (Figure S**2**; 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 was meticulously reported by Ribeirio *et al*.[80] 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** (Figure 4) and Scheme 7. Their observations and conclusions are: (1) during the reaction, structural and conformational fluctuations invoked by residues' electrostatic contributions lead to 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}[80] and very close to experimental value of 15.9 kcal mol^{-1}.[67] (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.[80]



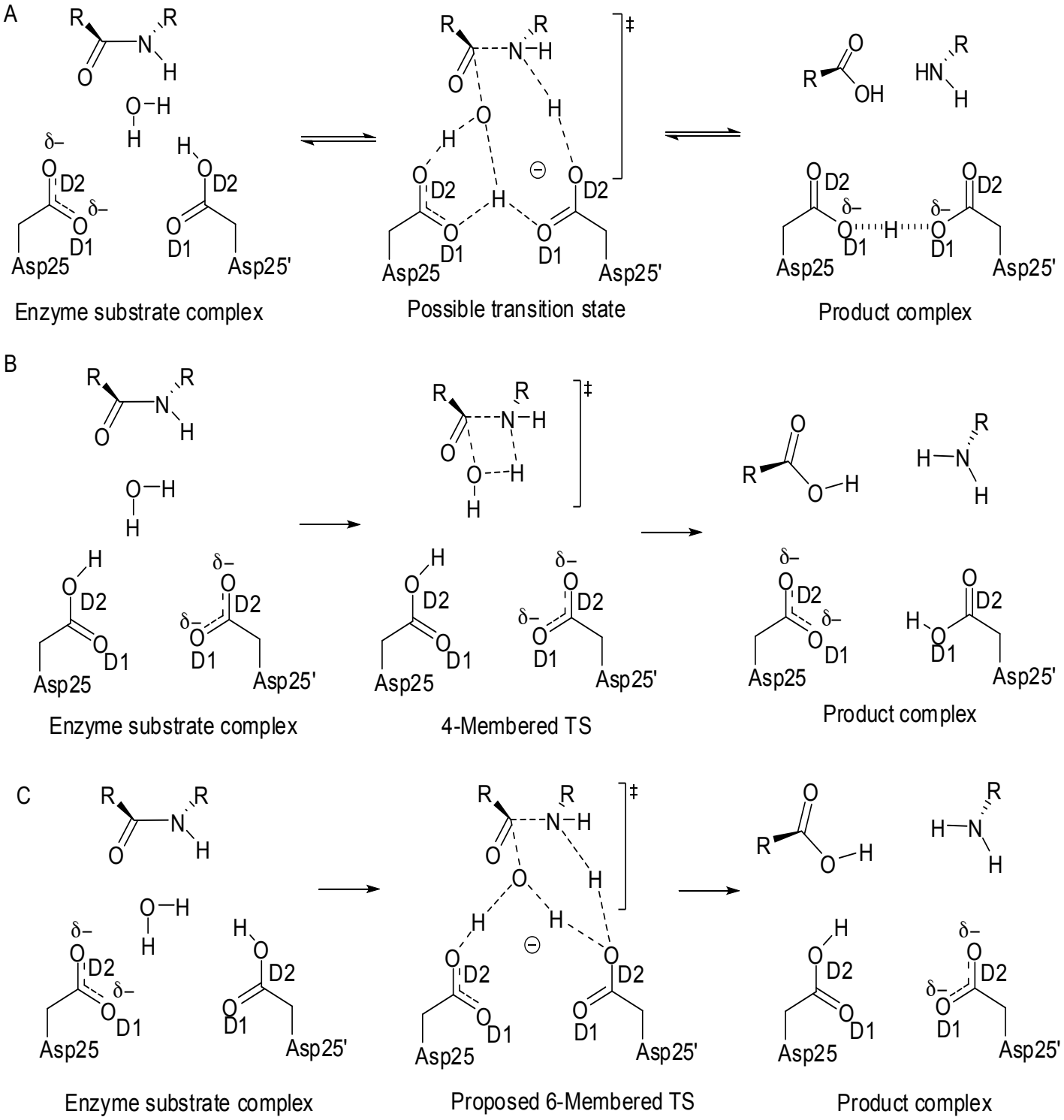
Scheme 7: Different proposed mechanism and configurations adopted by the active centre of protease redrawn from literature.[22, 23]. A; WatC nucleophilic attack and proton transfer to Asp25′, when it loses the planarity, the nucleophile can be more or less stabilized by the highlighted hydrogen (configurations 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.

The influence of frozen residues on the exploration of the PES of HIV-1 PR reaction mechanisms was recently studied.[165] The authors adopted Scheme 7 and previously reported computational protocols[80] 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 substrate.[165] 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,[165] an observation similar to investigations from here.[2, 18] 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.

4.3.2.2 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,[170] they have proposed a one-step catalytic model in which the nucleophilic WatC and the electrophilic aspartate (an acidic proton) 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.[170] The post-reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms[170] (model **J**; Figure 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[170] perspective. Note that the 6-membered ring TS in Scheme 8C is proposed by us.



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.*[24] hypothesis. (B) Concerted 4-membered ring model redrawn from literature,[25] the catalytic role of the aspartate groups seems apparently loss and (C) Proposed concerted 6-membered cyclic HIV-1 PR enzymatic mechanism.

In 2000, Park *et al.*[1] gave an investigation on the catalytic mechanism for Asp PRs using (MP2)[112] and HF[96] theories coupled with the 6-31G(2d)[97] 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 (Figure 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.[108] 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[108] 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[19, 148-154, 171] from our group whereby the 4-membered cyclic TS structures rarely yield better results than the 6-membered models (Scheme 8C).

5 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.

Based on investigations from both theoretical and experimental, HIV-1 PR recognizes its natural substrate through; substrate modulation, conserved substrate shape, interdependence conformational adaptability of both enzyme and targets.[9, 11, 31, 32] A notable contrasting opinion from computational simulations[36] 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.[36]

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 (Figure 10, INT1) is more widely accepted than the presence of the unstable oxyanion (Figure 10, INT2). Apart from 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[11, 24, 138, 172] and theoretical[81, 98, 173] techniques. The moieties along the enzyme—substrate reaction mechanism is 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.[81]

5.1 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[101, 131] and computational[17, 101, 106, 108, 131, 161] methods. A contrasting opinion[138] 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[80, 165] 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*.[113, 114] had earlier mentioned that it does not yield a discrete INT, therefore, the possibility of elucidating TS analogues 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.[170] In the light of literature support for similar six-membered ring transition structures observed theoretically,[79, 148-154] 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.

5.2 Future perspective

Studies on 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, 2, 17, 18, 95, 106, 131] is more popular compared to an entire representation of the enzyme—substrate system.[82, 98] 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, 2, 17, 18, 95, 106, 131]

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[10] has almost 200 wild type HIV-1 PR complexed with various ligands/substrate segments at resolution as low as 1.00 Å.[174] It is informative to note that high resolution does not necessarily mean the best starting geometry for computational simulation. For instance, 3NU3.pdb[175] was refined to 1.02 Å and the active site dihedral angel (^{Asp25}OD2--^{Asp25}OD1--$^{Asp25\prime}$OD1--$^{Asp25\prime}$OD2) is 49.43º, 4HVP[125] (refined to 2.30 Å) gave 15.57º for this angle, while as large as 67º have been noted for this important dihedral.[2] 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 exact protonation pattern of the Asp dyad is still a matter of debate,[87,90-93] 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.

Authors' information

Corresponding author

*E-mail: kruger@ukzn.ac.za

Competing interests

The authors declare no competing interests.

Acknowledgements

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.

References

[1] 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.

[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] 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*, pp 203-241, Springer.

[4] Lockhat, H. A., Silva, J. R., Alves, C. N., Govender, T., Lameira, J., Maguire, G. E., Sayed, Y., and Kruger, H. G. (2016) Binding Free Energy Calculations of Nine FDA-approved Protease Inhibitors Against HIV-1 Subtype C I36T↑ T Containing 100 Amino Acids Per Monomer, *Chemical Biology & Drug Design*.

[5] Maseko, S. B., Natarajan, S., Sharma, V., Bhattacharyya, N., Govender, T., Sayed, Y., Maguire, G. E., 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.

[6] Swanstrom, R., and Wills, J. (1997) *Synthesis, assembly, and processing of viral proteins*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

[7] 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.

[8] 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.

[9] 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 protease, *Journal of Molecular Biology 301*, 1207-1220.

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

[11] 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.

[12] Lv, Z., Chu, Y., and Wang, Y. (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity, *HIV/AIDS (Auckland, NZ) 7*, 95.

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

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

[15] Lee, S.-K., Potempa, M., Kolli, M., Özen, A., Schiffer, C. A., and Swanstrom, R. (2012) 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, *Journal of Biological Chemistry 287*, 13279-13290.
[16] Trylska, J. (2001) Computational Modelling of Protonation Equilibria and Reaction Mechanism of HIV-1 Protease.
[17] 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.
[18] Altoè, P., Stenta, M., Bottoni, A., and Garavelli, M. (2007) A tunable QM/MM approach to chemical reactivity, structure and physico-chemical properties prediction, *Theoretical Chemistry Accounts 118*, 219-240.
[19] Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design: Molecular Modeling, Spectroscopy, and Experimental Bioactivity, *Chemical Review 114*, 493-537.
[20] Gulnik, S., Erickson, J. W., and Xie, D. (2000) HIV protease: enzyme function and drug resistance, *Vitamins & Hormones 58*, 213-256.
[21] Rodriguez-Barrios, F., and Gago, F. (2004) HIV protease inhibition: Limited recent progress and advances in understanding current pitfalls, *Current Topics in Medicinal Chemistry 4*, 991-1007.
[22] Nijhuis, M., van Maarseveen, N. M., and Boucher, C. A. (2007) HIV protease resistance and viral fitness, *Current Opinion in HIV and AIDS 2*, 108-115.
[23] Brik, A., and Wong, C.-H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & Biomolecular Chemistry 1*, 5-14.
[24] Sussman, F., Villaverde, M. C., Dominguez, J. L., and Danielson, U. H. (2013) On the Active Site Protonation State in Aspartic Proteases: Implications for Drug Design, *Current Pharmaceutical Design 19*, 4257-4275.
[25] Deo, S. K., Lewis, J. C., and Daunert, S. (2000) Bioluminescence detection of proteolytic bond cleavage by using recombinant aequorin, *Analytical Biochemistry 281*, 87-94.
[26] Makatini, M. M., Petzold, K., Arvidsson, P. I., Honarparvar, B., Govender, T., Maguire, G. E. M., Parboosing, R., Sayed, Y., Soliman, M. E. S., 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.
[27] Pawar, S. A., Jabgunde, A. M., Govender, P., Maguire, G. E. M., Kruger, H. G., Parboosing, R., Soliman, M. E. S., 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.
[28] Makatini, M. M., Petzold, K., Alves, C. N., Arvidsson, P. I., Honarparvar, B., Govender, P., Govender, T., Kruger, H. G., Sayed, Y., Lameira, J., Maguire, G. E. M., and Soliman, M. E. S. (2013) Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors, *Journal of Enzyme Inhibition and Medicinal Chemistry 28*, 78-88.
[29] Debouck, C. (1992) The HIV-1 protease as a therapeutic target for AIDS, *AIDS Research and Human Retroviruses 8*, 153-164.
[30] 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.

[31] 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.
[32] Ozer, N., Haliloglu, T., and Schiffer, C. A. (2006) Substrate specificity in HIV-1 protease by a biased sequence search method, *Proteins: Structure, Function, and Bioinformatics 64*, 444-456.
[33] Ozer, N., Schiffer, C. A., and Haliloglu, T. (2006) Predicting substrate specificity in HIV-1 protease, pp 86-87, Computer Aided Drug Design & Development Society in Turkey.
[34] Ozen, A., Haliloglu, 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.
[35] Shen, Y., Altman, M. D., Ali, A., Nalam, M. N. L., Cao, H., Rana, T. M., Schiffer, C. A., and Tidor, B. (2013) Testing the substrate-envelope hypothesis with designed pairs of compounds, *ACS Chemical Biology 8*, 2433-2441.
[36] 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.
[37] 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.
[38] Tozser, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W., and Weber, I. T. (2000) Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases, *European Journal of Biochemistry 267*, 6287-6295.
[39] Jones, D. T., Tress, M., Bryson, K., and Hadley, C. (1999) Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure, *Proteins: Structure, Function, and Bioinformatics 37*, 104-111.
[40] Pitera, J. W., and Van Gunsteren, W. F. (2002) A comparison of non-bonded scaling approaches for free energy calculations, *Molecular Simulation 28*, 45-65.
[41] Shirts, M. R., Pitera, J. W., Swope, W. C., and Pande, V. S. (2003) Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins, *Journal of Chemical Physics 119*, 5740-5761.
[42] Blondel, A. (2004) Ensemble variance in free energy calculations by thermodynamic integration: Theory, optimal "Alchemical" path, and practical solutions, *Journal of Computational Chemistry 25*, 985-993.
[43] 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.
[44] Alder, B. J., and Wainwright, T. E. (1959) Studies in molecular dynamics. I. General method, *The Journal of Chemical Physics 31*, 459-466.
[45] Chellappan, S., Kiran Kumar Reddy, G., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Kairys, V., Fernandes, M. X., Altman, M. D., and Tidor, B. (2007) Design of Mutation-resistant HIV Protease Inhibitors with the Substrate Envelope Hypothesis, *Chemical Biology & Drug Design 69*, 298-313.
[46] Altman, M. D., Ali, A., Reddy, G. K. K., Nalam, M. N., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., and Gilson, M. K. (2008) HIV-1 protease inhibitors from

inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants, *Journal of the American Chemical Society 130*, 6099.

[47] Nalam, M. N., and Schiffer, C. A. (2008) New approaches to HIV protease inhibitor drug design II: testing the substrate envelope hypothesis to avoid drug resistance and discover robust inhibitors, *Current Opinion in HIV and AIDS 3*, 642.

[48] Nalam, M. N., Ali, A., Altman, M. D., Reddy, G. K. K., Chellappan, S., Kairys, V., Özen, A., Cao, H., Gilson, M. K., and Tidor, B. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance, *Journal of Virology 84*, 5368-5378.

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

[50] Dunbrack, R. L., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences, *Protein Science 6*, 1661-1681.

[51] Pettit, S. C., Henderson, G. J., Schiffer, C. A., and Swanstrom, R. (2002) 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, *Journal of Virology 76*, 10226-10233.

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

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

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

[55] Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein Folding and Association - Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons, *Proteins 11*, 281-296.

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

[57] Özen, A., Haliloğlu, T., and Schiffer, C. A. (2012) HIV-1 protease and substrate coevolution validates the substrate envelope as the substrate recognition pattern, *Journal of Chemical Theory and Computation 8*.

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

[59] Pai, V. B., and Nahata, M. C. (1999) Nelfinavir mesylate: a protease inhibitor, *Annals of Pharmacotherapy 33*, 325-339.

[60] Özer, N., Özen, A., Schiffer, C. A., and Haliloğlu, T. (2015) Drug-resistant HIV-1 protease regains functional dynamics through cleavage site coevolution, *Evolutionary Applications 8*, 185-198.

[61] Atilgan, A. R., Durell, S., Jernigan, R. L., Demirel, M., Keskin, O., and Bahar, I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophysical Journal 80*, 505-515.
[62] Laco, G. S. (2015) HIV-1 protease substrate-groove: role in substrate recognition and inhibitor resistance, *Biochimie 118*, 90-103.
[63] Wang, Y., Dewdney, T. G., Liu, Z., Reiter, S. J., Brunzelle, J. S., Kovari, I. A., and Kovari, L. C. (2012) Higher desolvation energy reduces molecular recognition in multi-drug resistant HIV-1 protease, *Biology 1*, 81-93.
[64] Liu, Z., Wang, Y., Brunzelle, J., Kovari, I. A., and Kovari, L. C. (2011) Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease, *The Protein Journal 30*, 173-183.
[65] Yedidi, R. S., Proteasa, G., Martinez, J. L., Vickrey, J. F., Martin, P. D., Wawrzak, Z., Liu, Z., Kovari, I. A., and Kovari, L. C. (2011) Contribution of the 80s loop of HIV-1 protease to the multidrug-resistance mechanism: crystallographic study of MDR769 HIV-1 protease variants, *Acta Crystallographica Section D: Biological Crystallography 67*, 524-532.
[66] Liu, Z., Wang, Y., Yedidi, R. S., Dewdney, T. G., Reiter, S. J., Brunzelle, J. S., Kovari, I. A., and Kovari, L. C. (2013) Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes, *Biochemical and Biophysical Research Communications 430*, 1022-1027.
[67] Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and 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*, 33231-33235.
[68] Tozser, J., Bagossi, P., Weber, I. T., Copeland, T. D., and Oroszlan, S. (1996) Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases, *Journal of Biological Chemistry 271*, 6781-6788.
[69] Feher, A., Weber, I. T., Bagossi, P., Boross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Torshin, I. Y., Harrison, R. W., and Tozser, J. (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites, *European Journal of Biochemistry 269*, 4114-4120.
[70] Tozser, J., Gustchina, A., Weber, I. T., Blaha, I., Wondrak, E. M., and Oroszlan, S. (1991) Studies on the Role of the S4 Substrate Binding-Site of Hiv Proteinases, *Febs Letters 279*, 356-360.
[71] Altman, M. D., Nalivaika, E. A., Prabu-Jeyabalan, M., Schiffer, C. A., and Tidor, B. (2008) Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease, *Proteins-Structure Function and Bioinformatics 70*, 678-694.
[72] Tripathi, A., Fornabaio, M., Spyrakis, F., Mozzarelli, A., Cozzini, P., and Kellogg, G. E. (2007) Complexity in modeling and understanding protonation states: Computational titration of HIV-1-protease-inhibitor complexes, *Chemistry & Biodiversity 4*, 2564-2577.
[73] Weber, I. T., Waltman, M. J., Mustyakimov, M., Blakeley, M. P., Keen, D. A., Ghosh, A. K., Langan, P., and Kovalevsky, A. Y. (2013) Joint X-ray/neutron crystallographic study of HIV-1 protease with clinical inhibitor amprenavir: insights for drug design, *Journal of Medicinal Chemistry 56*, 5631-5635.
[74] Gerlits, O., Wymore, T., Das, A., Shen, C. H., Parks, J. M., Smith, J. C., Weiss, K. L., Keen, D. A., Blakeley, M. P., and Louis, J. M. (2016) Long-Range Electrostatics-Induced Two-

Proton Transfer Captured by Neutron Crystallography in an Enzyme Catalytic Site, *Angewandte Chemie International Edition 55*, 4924-4927.

[75] Yamazaki, T., Nicholson, L. K., Wingfield, P., Stahl, S. J., Kaufman, J. D., Eyermann, C. J., Hodge, C. N., Lam, P. Y., and Torchia, D. (1994) 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, *Journal of the American Chemical Society 116*, 10791-10792.

[76] Smith, R., Brereton, I. M., Chai, R. Y., and Kent, S. B. H. (1996) Ionization states of the catalytic residues in HIV-1 protease, *Nature Structural Biology 3*, 946-950.

[77] Wang, Y. X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., and 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*, 9945-9950.

[78] Yu, N., Hayik, S. A., Wang, B., Liao, N., Reynolds, C. H., and Merz Jr, K. M. (2006) Assigning the protonation states of the key aspartates in β-secretase using QM/MM X-ray structure refinement, *Journal of Chemical Theory and Computation 2*, 1057.

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

[80] Ribeiro, A. J. M., Santos-Martins, D., Russo, N., Rarnos, M. J., and Fernandes, P. A. (2015) Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease, *ACS Catalysis 5*, 5617-5626.

[81] 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.

[82] Okimoto, N., Kitayama, K., Hata, M., Hoshino, T., and Tsuda, M. (2001) Molecular dynamics simulations of a complex of HIV-1 protease and substrate: substrate-dependent efficiency of catalytic activity, *Journal of Molecular Structure: Theochem 543*, 53-63.

[83] Itoh, S. G., Damjanovic, A., and Brooks, B. R. (2012) pH Replica-Exchange Method Based on Discrete Protonation States, *Biophysical Journal 102*, 40a-40a.

[84] Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Structure, Function, and Bioinformatics 61*, 704-721.

[85] Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic Acids Research 35*, W522-W525.

[86] Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating p K as and adding missing hydrogens to macromolecules, *Nucleic Acids Research 33*, W368-W371.

[87] Mcgee, T. D., Edwards, J., and Roitberg, A. E. (2014) pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State, *The Journal of Physical Chemistry B 118*, 12577-12585.

[88] Piana, S., and 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*, 26-36.
[89] Dodson, G., and Wlodawer, A. (1998) Catalytic triads and their relatives, *Trends in Biochemical Science 23*, 347-352.
[90] Makatini, M. M., Petzold, K., Sriharsha, S. N., Soliman, M. E. S., Honarparvar, B., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Pentacycloundecane-based inhibitors of wild-type C-South African HIV-protease, *Bioorganic and Medicinal Chemistry Letters 21*, 2274-2277.
[91] 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., and Debouck, C. (1989) Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease, *Proc. Natl. Acad. Sci. U. S. A. 86*, 1841-1845.
[92] Smith, R., Brereton, I. M., Chai, R. Y., and Kent, S. B. (1996) Ionization states of the catalytic residues in HIV-1 protease, *Nature Structural & Molecular Biology 3*, 946-950.
[93] Piana, S., Sebastiani, D., Carloni, P., and 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*, 8730-8737.
[94] Piana, S., Carloni, P., and Rothlisberger, U. (2002) Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations, *Protein Science 11*, 2393-2402.
[95] Piana, S., Carloni, P., and Parrinello, M. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease, *Journal of Molecular Biology 319*, 567-583.
[96] Becke, A. D. (1993) A new mixing of Hartree–Fock and local density-functional theories, *The Journal of Chemical Physics 98*, 1372-1377.
[97] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G* basis set for third-row atoms, *Journal of Computational Chemistry 22*, 976-984.
[98] Bjelic, S., and Aqvist, J. (2006) Catalysis and linear free energy relationships in aspartic proteases, *Biochemistry 45*, 7709-7723.
[99] 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.
[100] 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.
[101] Rodriguez, E. J., Angeles, T. S., and Meek, T. D. (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.
[102] Lee, H., Darden, T. A., and Pedersen, L. G. (1996) An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease, *Journal of the American Chemical Society 118*, 3946-3950.
[103] Liu, H. Y., MullerPlathe, F., and vanGunsteren, W. F. (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease, *Journal of Molecular Biology 261*, 454-469.
[104] 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.

[105] Okimoto, N., Makiyama, M., Hata, M., and Tsuda, M. (1997) Gag protein hydrolysis mechanism by HIV-1 protease - Investigation by semiempirical molecular orbital method, *Nippon Kagaku Kaishi*, 260-266.
[106] Okimoto, N., Tsukui, T., Hata, M., Hoshino, T., and Tsuda, M. (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method, *Journal of the American Chemical Society 121*, 7349-7354.
[107] Trylska, J., Bala, P., Geller, M., and Grochowski, P. (2002) Molecular dynamics simulations of the first steps of the reaction catalyzed by HIV-1 protease, *Biophys J 83*, 794-807.
[108] 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.
[109] Grochowski, P., Lesyng, B., Bała, P., and McCammon, J. (1996) Density functional based parametrization of a valence bond method and its applications in quantum-classical molecular dynamics simulations of enzymatic reactions, *International Journal of Quantum Chemistry 60*, 1143-1164.
[110] Trylska, J., Grochowski, P., and Geller, M. (2001) Parameterization of the approximate valence bond (AVB) method to describe potential energy surface in the reaction catalyzed by HIV-1 protease, *International Journal of Quantum Chemistry 82*, 86-103.
[111] Hohenberg, P., and Kohn, W. (1964) Inhomogeneous electron gas, *Physical Review 136*, B864.
[112] Møller, C., and Plesset, M. S. (1934) Note on an approximation treatment for many-electron systems, *Physical Review 46*, 618.
[113] Hyland, L. J., Tomaszek Jr, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., and 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.
[114] 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.
[115] Antonov, V., Ginodman, L., Kapitannikov, Y. V., Barshevskaya, T., Gurova, A., and Rumsh, L. (1978) Mechanism of pepsin catalysis: General base catalysis by the active-site carboxylate ion, *Febs Lett 88*, 87-90.
[116] Antonov, V. K., Ginodman, L. M., Rumsh, L. D., Kapitannikov, Y. V., Barshevskaya, T. N., Yavashev, L. P., Gurova, A. G., and Volkova, L. I. (1981) Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange, *The FEBS Journal 117*, 195-200.
[117] 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.
[118] 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.
[119] Kästner, J. (2011) Umbrella sampling, *Wiley Interdisciplinary Reviews: Computational Molecular Science 1*, 932-942.

[120] Laio, A., VandeVondele, J., and Rothlisberger, U. (2002) A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations, *The Journal of Chemical Physics 116*, 6941-6947.
[121] Becke, A. D. (1988) Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior, *Physical Review A 38*, 3098-3100.
[122] 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.
[123] Kresse, G., and Furthmüller, J. (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, *Computational Materials Science 6*, 15-50.
[124] 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.
[125] Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. (1989) Structure of Complex of Synthetic HIV-1 Protease with a Substrate-Based Inhibitor At 2.3-A Resolution, *Science 246*, 1149-1152.
[126] Schock, H. B., Garsky, V. M., and Kuo, L. C. (1996) 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 271*, 31957-31963.
[127] Aqvist, J., and Warshel, A. (1993) Simulation of Enzyme-Reactions using Valence-Bond Force-Fields and other Hybrid Quantum-Classical Approaches, *Chem Rev 93*, 2523-2544.
[128] Kamerlin, S. C. L., and Warshel, A. (2011) The empirical valence bond model: theory and applications, *Wires Comput Mol Sci 1*, 30-45.
[129] Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, Wiley New York.
[130] Cascella, M., Micheletti, C., Rothlisberger, U., and Carloni, P. (2005) Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases, *Journal of the American Chemical Society 127*, 3734-3742.
[131] 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.
[132] Schramm, V. L. (2013) Transition states, analogues and drug development, *ACS Chemical Biology 8*, 71.
[133] Kipp, D. R., Silva, R. G., and Schramm, V. L. (2011) Mass-Dependent Bond Vibrational Dynamics Influence Catalysis by HIV-1 Protease, *Journal of the American Chemical Society 133*, 19358-19361.
[134] Torbeev, V. Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W. M., Perozo, E., and Kent, S. B. H. (2011) Protein conformational dynamics in the mechanism of HIV-1 protease catalysis, *P Natl Acad Sci USA 108*, 20982-20987.
[135] Polgar, L., Szeltner, Z., and Boros, I. (1994) Substrate-dependent mechanisms in the catalysis of human immunodeficiency virus protease, *Biochemistry 33*, 9351-9357.
[136] Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A 38*, 3098.
[137] Kovalevsky, A. Y., Chumanevich, A. A., Liu, F., Louis, J. M., and Weber, I. T. (2007) Caught in the act: The 1.5 angstrom resolution crystal structures of the HIV-1 protease and the 154V mutant reveal a tetrahedral reaction lntermediate, *Biochemistry 46*, 14854-14864.

[138] 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.
[139] Darden, T., Perera, L., Li, L., and Pedersen, L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations, *Structure (London, England) 7*, R55-R60.
[140] 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.
[141] Binkley, J. S., Pople, J. A., and Hehre, W. J. (1980) Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements, *Journal of the American Chemical Society 102*, 939-947.
[142] Gordon, M. S., Binkley, J. S., Pople, J. A., Pietro, W. J., and Hehre, W. J. (1982) Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements, *Journal of the American Chemical Society 104*, 2797-2803.
[143] Hehre, W. J., Stewart, R. F., and Pople, J. A. (1969) self-consistent molecular-orbital methods. i. use of gaussian expansions of Slater-type atomic orbitals, *The Journal of Chemical Physics 51*, 2657-2664.
[144] Collins, J. B., von R. Schleyer, P., Binkley, J. S., and Pople, J. A. (1976) Self-consistent molecular orbital methods. XVII. Geometries and binding energies of second-row molecules. A comparison of three basis sets, *The Journal of Chemical Physics 64*, 5142-5151.
[145] 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, *J Comput Chem 24*, 1999-2012.
[146] Gerritzen, D., Limba, and , H.-H. (1984) NMR Spectroscopy, *Journal of the American Chemical Society 106*, 869-879.
[147] Cleland, W. (2000) Low-barrier hydrogen bonds and enzymatic catalysis, *Archives of Biochemistry and Biophysics 382*, 1-5.
[148] Kruger, H. G. (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *Journal of Molecular Structure: Theochem 577*, 281-285.
[149] 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[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam, *J Mol Struc-Theochem 672*, 119-125.
[150] 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, *J Mol Struc-Theochem 771*, 165-170.
[151] 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.
[152] 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, *Org Biomol Chem 13*, 9993-10006.

[153] Lawal, M. M., Govender, T., Maguire, G. E., 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.
[154] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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, *J Mol Struct 1128*, 94-102.
[155] Rose, R. B., Craik, C. S., Douglas, N. L., and Stroud, R. M. (1996) Three-Dimensional Structures of HIV-1 and SIV Protease Product Complexes†, *Biochemistry 35*, 12933-12944.
[156] 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.
[157] 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., and Pannu, N. S. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallographica Section D: Biological Crystallography 54*, 905-921.
[158] 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.
[159] 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.
[160] 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*, e7860.
[161] 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.
[162] Ichikawa, M. (1979) The C–O vs O–H length correlation in hydrogen-bonded carboxyl groups, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry 35*, 1300-1301.
[163] Yu, N., Hayik, S. A., Wang, B., Liao, N., Reynolds, C. H., and Merz, K. M. (2006) Assigning the protonation states of the key aspartates in beta-secretase using QM/MM X-ray structure refinement, *J Chem Theory Comput 2*, 1057-1069.
[164] Weber, I. T., Agniswamy, J., Fu, G. X., Shen, C. H., and Harrison, R. W. (2012) Reaction Intermediates Discovered in Crystal Structures of Enzymes, In *Structural and Mechanistic Enzymology: Bringing Together Experiments and Computing* (Christov, C., and KarabenchevaChristova, T., Eds.), pp 57-86.
[165] Calixto, A. R., Ramos, M. J., and Fernandes, P. A. (2017) Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms, *J Chem Theory Comput 13*, 5486-5495.
[166] Mahalingam, A. K., Axelsson, L., Ekegren, J. K., Wannberg, J., Kihlström, J., Unge, T., Wallberg, H., Samuelsson, B., Larhed, M., and Hallberg, A. (2009) HIV-1 protease

inhibitors with a transition-state mimic comprising a tertiary alcohol: improved antiviral activity in cells, *Journal of medicinal chemistry 53*, 607-615.
[167] Carnevale, V., Raugei, S., Piana, S., and Carloni, P. (2008) On the nature of the reaction intermediate in the HIV-1 protease: a quantum chemical study, *Comput Phys Commun 179*, 120-123.
[168] 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.
[169] 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.
[170] 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.
[171] Lawal, M. M., Govender, T., Maguire, G. E., Kruger, H. G., and Honarparvar, B. (2017) 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.
[172] Wang, Y.-X., Freedberg, D. I., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., Bhat, T. N., Erickson, J. W., and Torchia, D. A. (1996) Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR, *Journal of the American Chemical Society 118*, 12287-12290.
[173] Li, Z., and Lazaridis, T. (2003) Thermodynamic contributions of the ordered water molecule in HIV-1 protease, *Journal of the American Chemical Society 125*, 6636-6637.
[174] Ghosh, A. K., Gemma, S., Baldridge, A., Wang, Y. F., Kovalevsky, A. Y., Koh, Y., Weber, I. T., and Mitsuya, H. (2008) Flexible cyclic ethers/polyethers as novel P2-ligands for HIV-1 protease inhibitors: Design, synthesis, biological evaluation, and protein-ligand X-ray studies, *Journal of Medicinal Chemistry 51*, 6021-6033.
[175] Shen, C. H., Wang, Y. F., Kovalevsky, A. Y., Harrison, R. W., and Weber, I. T. (2010) Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters, *Febs Journal 277*, 3699-3714.

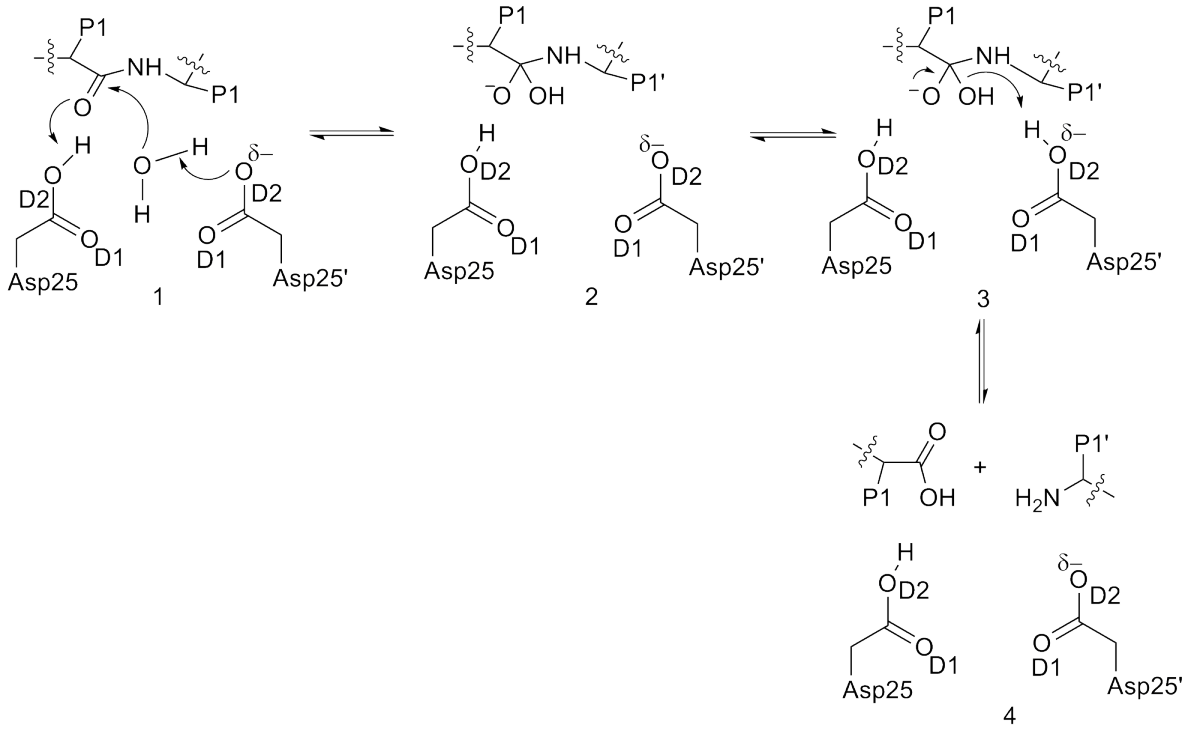
Supporting Information for Chapter two

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

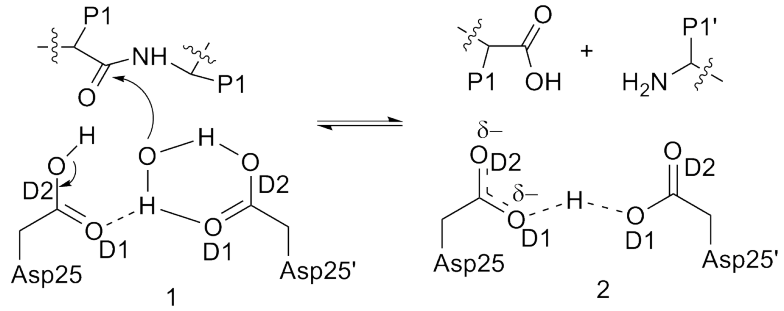
Monsurat M. Lawal,[a] Zainab K. Sanusi,[a] Thavendran Govender,[a] Glenn E. M. Maguire,[a, b] Bahareh Honarparvar,[a] and Hendrik G. Kruger.[a,*]

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

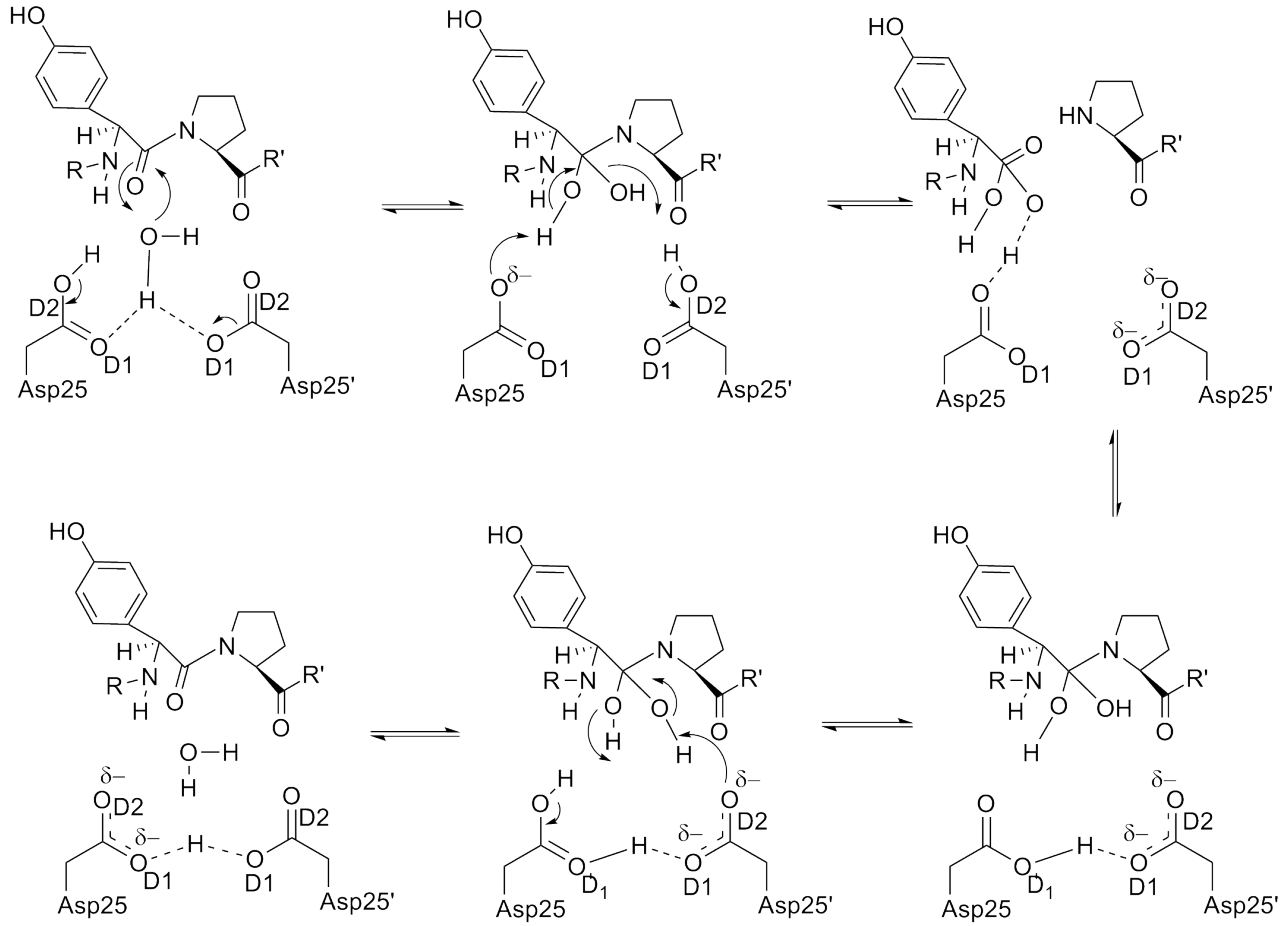
[b]School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4041, South Africa.



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.



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.

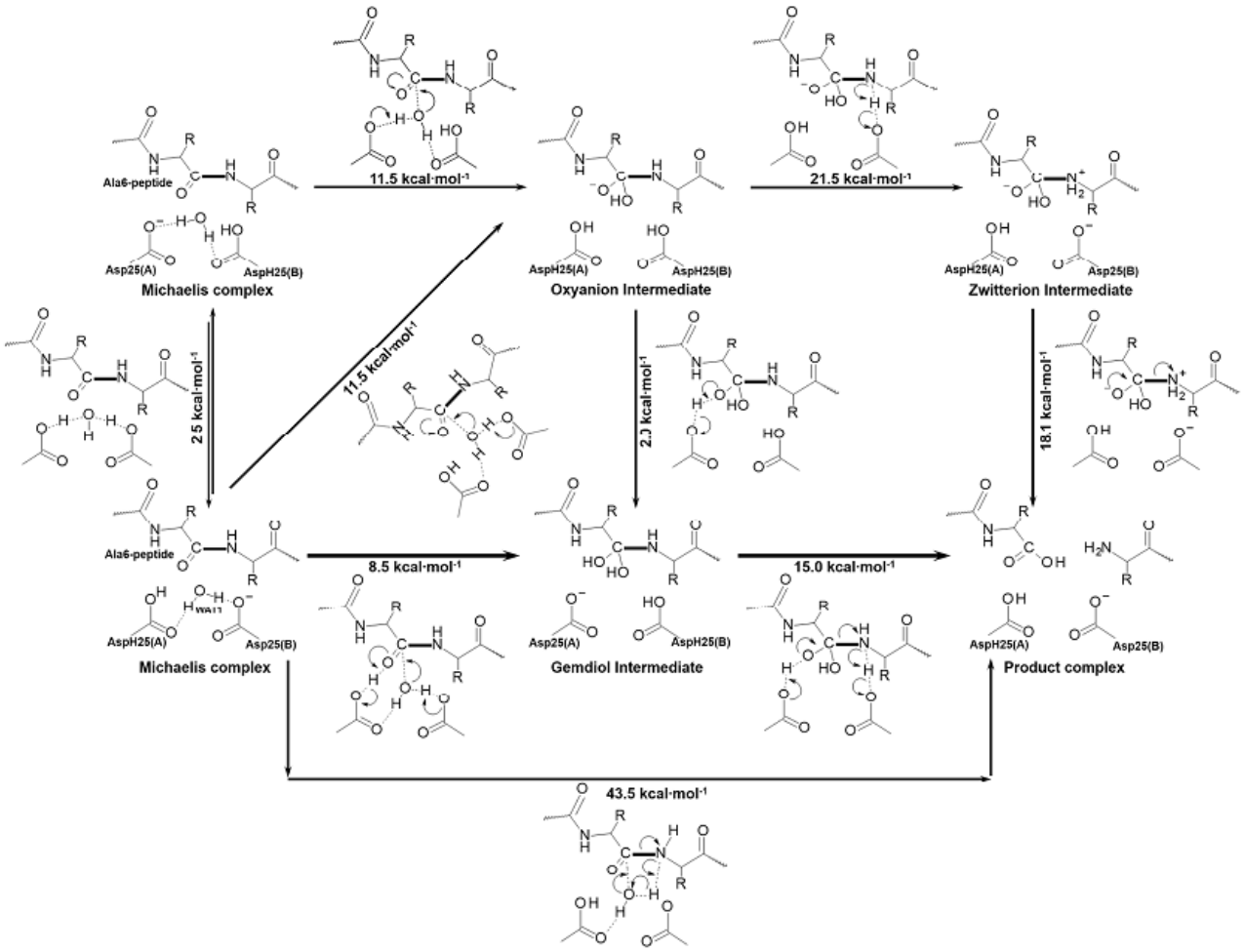


Figure S1: Schematic representation of the overall studied mechanisms for HIV-1 PR catalysed reaction of hexapeptide bond cleavage. Activation free energies, in relation to the initial reactant state, were derived from the AM1/MM PMFs corrected at M06-2X:AM1/MM level. All values are reported in kcal·mol-1.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

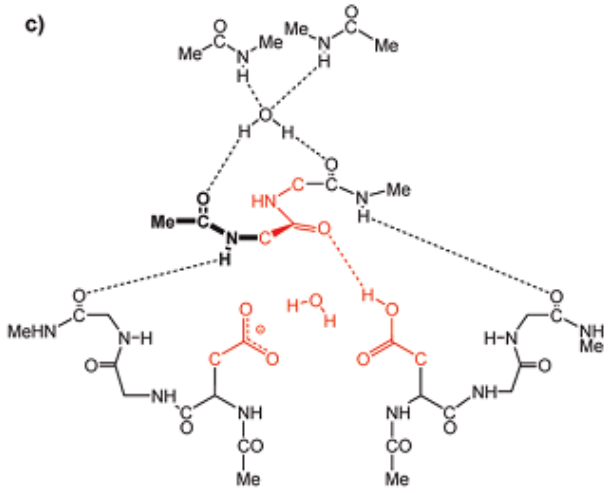


Figure S2: In ONIOM calculations, atoms in red are described at the high level.[2]

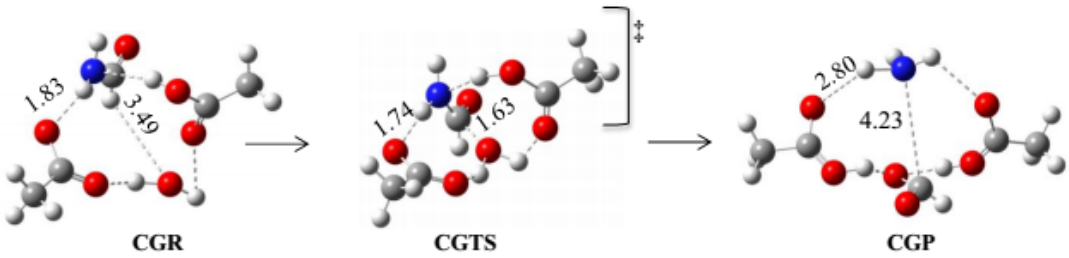


Figure S3: The concerted general acid-base mechanism involving 8-membered ring TS.[3]

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

H-Bond	INT	TS	Prod
O_{WatC}—$H1_{WatC}$--$OD1_{Asp25}$	2.30[4]		2.65[6]
	2.70[5]		2.90[5]
OD1—H_{Asp25}'--N	2.50[4]		
$OD2_{Asp25}$--O_{WatC}	2.50[7]	2.56[9]	2.45[6]
	2.91[8]		2.80[5]
	3.10[5]		
$OD1_{Asp25}$--$OD1_{Asp25}$'	2.60[9]	2.98[9]	2.30[10]
	2.48[7]		2.76[6]
	2.80[5]		
$OD2_{Asp25}$'--N	2.80/2.90[7]		2.90[10]
	3.10[5]		
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.

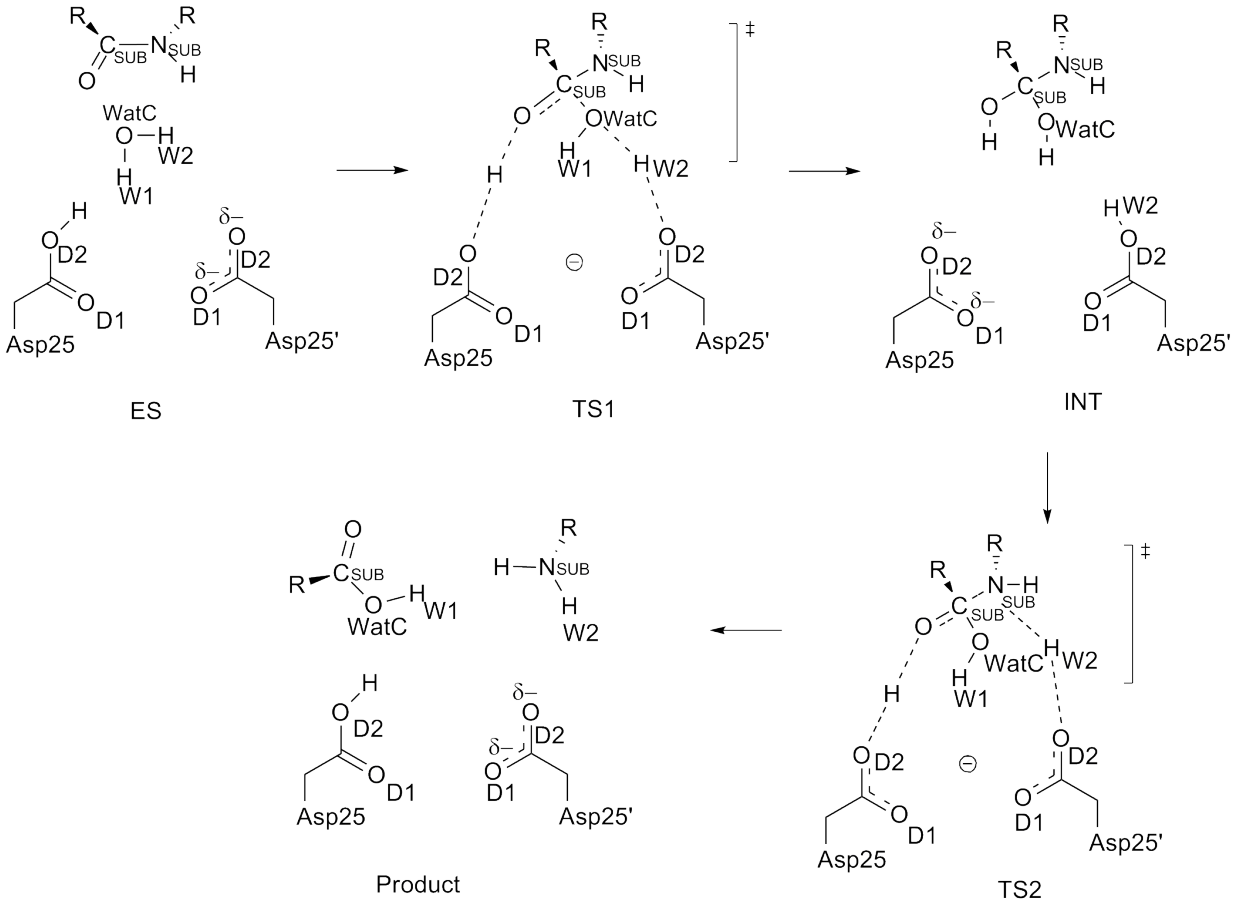


Figure S4: Complexes along mechanistic pathway of HIV-1 PR—substrates system.

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.

CHAPTER THREE

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

Monsurat M. Lawal,[a] Zainab K. Sanusi,[a] Collins U. Ibeji,[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.

Abstract

The catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease (PR) is one of the most studied aspartate protease representatives. Both experimental and theoretical techniques have been harnessed to provide a profound understanding of several possible reaction pathways for HIV-1 PR catalysis of the natural substrate/ligand. Most of these studies have investigated the stepwise general acid-base mechanism. The hydrolytic reaction has been reported to be viewed as a one-step process; the nucleophile (water molecule) and electrophile (an acidic proton) attack the scissile bond in a concerted manner, but little attention has been paid to this synchronous model. Herein, the one-step concerted catalytic mechanism of HIV-1 PR on its natural substrate and N-substituted derivatives were studied using density functional theory (DFT) method. The reactions proceed through the formation of a concerted transition state structure leading to product complexes. The applied *in silico* method allows the elucidation of activation parameters, kinetics, and quantum chemical properties for the systems. Using this model, we then examined the correlation between the adiabatic stretching force constants and the transition state energies of the system, which reflects bond strength and chemical reactivity. This model may allow the design of potential new natural substrate-based inhibitors by controlling the scissile peptide bond strength *via* electronic fine-tuning.

Keywords: catalysis; HIV-1 PR; concerted mechanism; density functional theory (DFT); scissile bond strength; transition state modelling.

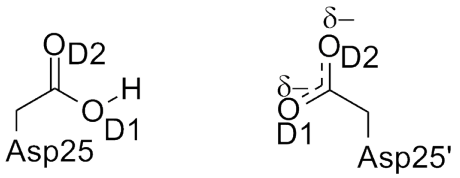
1 Introduction

HIV-1 PR is a crucial enzyme in the life cycle of the HIV-1 virus; the enzyme performs two distinct roles, which can be denoted as recognition and catalysis.[1, 2] Despite its unique homodimeric nature, HIV-1 PR recognizes its asymmetric natural substrates.[2] Recognition of substrate/ligand by this enzyme has been proposed to be achieved through substrate modulation, conserved substrate shape, interdependence conformational adaptability of both enzyme and targets[2-5] as well as conformational specificity of the protease to the Gag and Gag-Pol polyprotein pool.[6] Immature virion proteins encoded by the Gag and Gag-Pol polyproteins precursors have natural substrate sequences; capsid (CA), matrix (MA), spacer peptide 1 (p1), nucleocapsid (NC), spacer peptide 2 (p2), p6gag,[7] reverse transcriptase (RT), RNAse H (RH), and integrase (IN).[8] When HIV-1 PR cleaves these sites, nine cleavage domains are produced which is thereafter hydrolysed by this same protease.[2, 8]

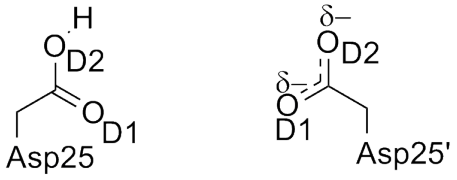
The nature of the hydrolytic mechanism of HIV-1 PR on its natural substrate is still undergoing intense debate. Earlier experimental studies on the catalytic mechanism focused on kinetics,[9-13] proposed mechanistic pathways[11, 14] and the nature of the reaction.[12, 15, 16] Recent experimental advances employed in trying to understand the catalytic HIV-1 PR—substrate process has involved the use of joint X-ray/neutron crystallographic technique to capture reaction complexes.[17-23] Earlier attempts to observe reacting moieties experimentally were futile.[17, 24] Computational techniques had long been embraced to elucidate the mechanistic landscape of enzymatic processes.[25, 26] More recently, the use of advanced theoretical tools to observe possible reaction complexes along the potential energy surface (PES) of HIV-1 PR—substrate was investigated.[1]

From both experimental and theoretical studies, the enzymatic mechanism of this reaction has been widely investigated as a stepwise general acid-base mechanism involving two[24, 26-31] and three[1, 11, 32] transition state (TS) structures. This process involves a catalytic water (WatC) in addition to the HIV-1 PR—substrate complex and the hydrolysis occurs in a number of steps (Scheme 1) in which the characteristic aspartate dyad of the PR is often monoprotonated (Figure 1). Another proposed mechanistic pathway is the nucleophilic process,[24, 26, 33] which is a variance

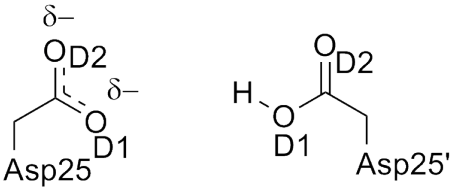
of the previous mechanism. The cleavage of natural substrate was thus widely proposed to operate as either one of these two mechanisms.[1, 11, 24, 26-33]



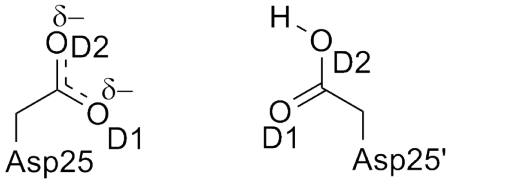
Model A; OD1 of Asp25 is protonated



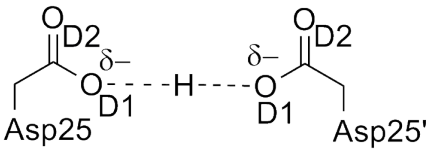
Model B; OD2 of Asp25 is protonated



Model C; OD1 of Asp25' is protonated



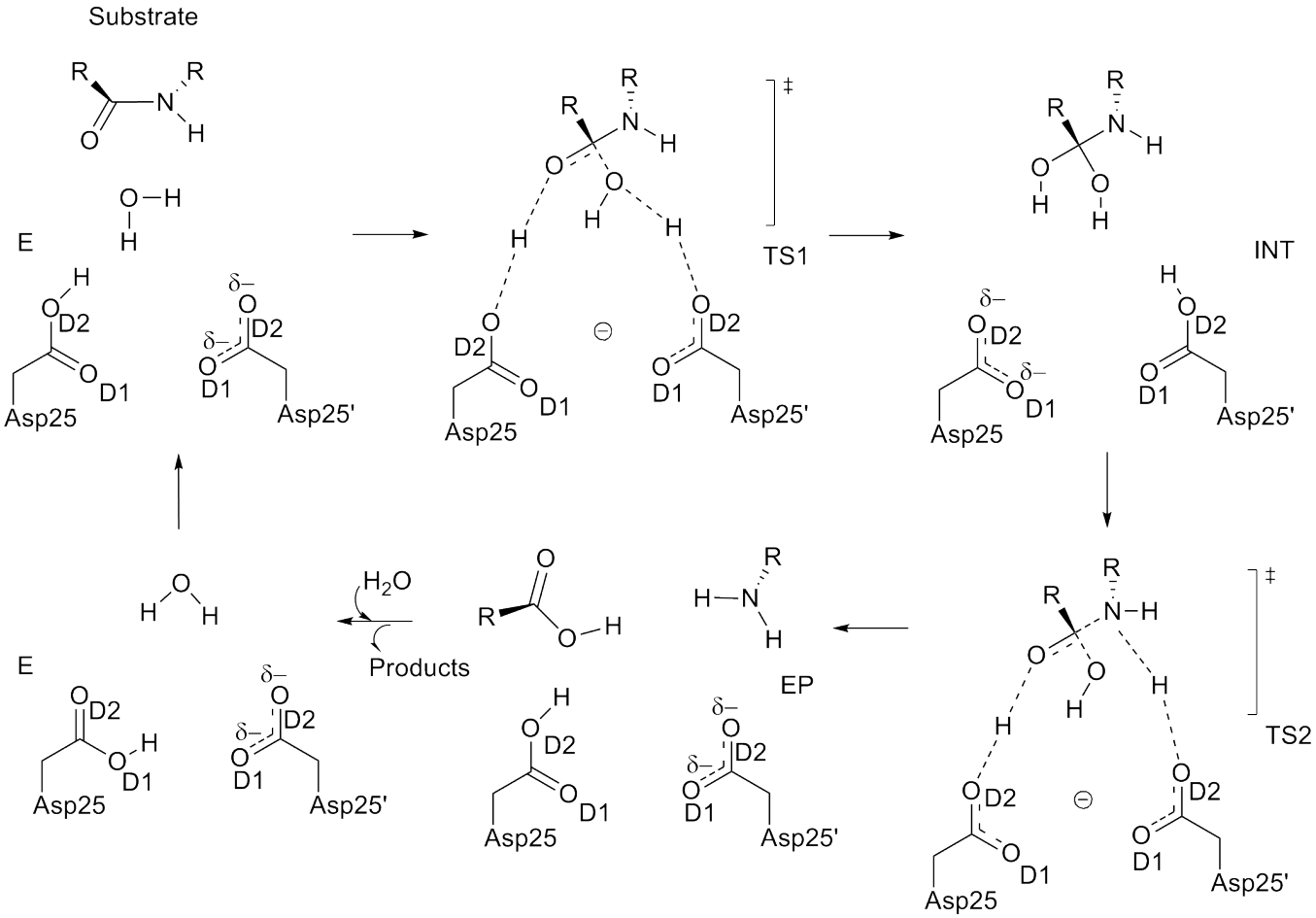
Model D; OD2 of Asp25' is protonated



Model E; Low barrier hydrogen stabilizes OD1 of Asp25 and Asp25'

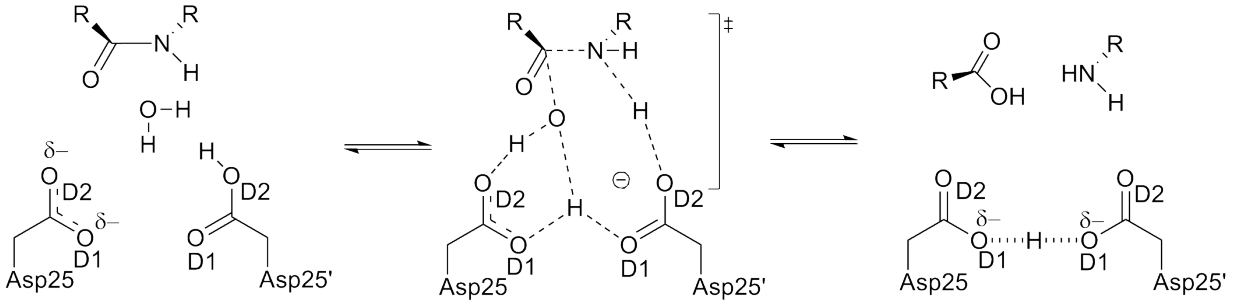
Figure 1: Possible monoprotonation models for the aspartate dyad.

Despite a vast number of both experimental and theoretical investigations on HIV-1 PR catalysis, some research questions are yet to be fully attended to. Amongst these is the theme of this present study, which are: (1) what is the feasibility of the earlier proposed[14] concerted general acid-base mechanism for this system? (2) What would the TS structure look like?



Scheme 1: Stepwise general acid-base catalytic mechanism of HIV-1 PR.[34] ES is the enzyme—substrate complex, TS 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. Adopted from literature.[34]

In 1991, Jaskólski *et al.* from experimental results, proposed a one-step hydrolytic mechanism for HIV-1 PR and its substrate.[14] They suggested a catalytic model in which the nucleophilic WatC and the electrophilic aspartate (an acidic proton) attack the scissile bond in a synchronous manner. At the beginning of the reaction, the acidic proton is located on the OD2 atom of the aspartate that is proximal to the N atom of the approaching amide.[14] The post reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms[14] (Model E; Figure 1) The proposed experimental protocol is summarized in Scheme 2. It is quite puzzling to observe that only two theoretical research outputs are found with respect to this concerted general acid-base mechanism.



Scheme 2: Proposed concerted HIV-1 PR mechanism based on Jaskólski et al.[14] hypothesis.

In 2000, Park *et al*.[24] reported an investigation on a model for the catalytic mechanism of Asp PRs. An extensive *ab initio* computation was carried out using Møller-Plesset second-order perturbation (MP2)[35] and Hartree-Fock (HF)[36] theories coupled with the 6-31G(2d)[37] basis set. They used a model reaction where formamide represents the substrate, while acetate and acetic acid represent the catalytic Asp residues.[24] This enabled these authors[24] to investigate a concerted nucleophilic (CN), stepwise nucleophilic (SN), stepwise general acid-base (SG) and concerted general acid-base (CG) mechanisms. The latter (CG) model gave an activation free energy of 30 kcal mol^{-1} and featured an 8-membered ring TS structure (Figure S1; supporting information, SI).

The latest attempt on the concerted catalytic mechanism of HIV-1 with a peptide substrate was presented by Krzemińska *et al*. in 2016.[1] The study utilizes an intriguing model in which both electrostatic and dynamic contributions were accounted for. The applied theoretical methods (umbrella sampling method; QM/MM/MD) enabled these authors[1] to explore a 4-membered ring TS model devoid of the catalytic (or general acid-base) function of the aspartic moieties (Scheme 1A; SI) with a substantial Gibb's free energy barrier (43.5 kcal mol^{-1}). This theoretical observation is quite possible, as in related chemical and enzymatic reaction[38-46] results from our group. The 4-membered cyclic TS structures rarely yield better results than the 6-membered models (Scheme 1B; SI).

The reader may be curious as to the reason for revisiting the concerted general acid-base mechanism, since the calculated activation free energies of 30[24] and 43.5[1] kcal mol^{-1} for the concerted hydrolysis of peptides by HIV-1 PR are higher than estimated experimental activation free energy within 14.86 and 21.03 kcal mol^{-1}.[9, 10, 47-52] The answer to this can be traced to the inaccuracy of previous attempts on the concerted mechanism reported by Park *et al*.[24] and

Krzemińska *et al*.[1] These results could be due to the computational protocol employed and the studied TS model, respectively, in each study.

In light of literature support for similar 6-membered ring transition structures observed theoretically,[38-41, 43-45, 53] a one-step concerted chemical process appears plausible and can potentially provide a new understanding of the catalytic mechanism of HIV-1 PR. Such a computational investigation appears absent and would be quite informative if the possibility of a single rate-determining step can be achieved. The suitability of DFT methods to elucidate mechanistic pathways and obtain data in broad agreement with experimental results is quite appealing.[44-46, 54] Specifically, Garrec *et al*.[54] examined the reactivity of an HIV-1 PR model to test the accuracy of the latest DFT methods in 2011. They[54] have suggested B3LYP[55]/6-31++G(d,p)[37] or B3LYP/6-311++G(d,p) combination for geometry optimization and MP2[35]/6-31+G(d,p) or MP2/6-31++G(d,p) model for single-point calculation. The accuracy of M06-2X[56]/def2-TZVP[57, 58] combinations for geometry optimization and thermochemistry prediction in esterification reactions of acid halides and acetic acid was recently reported from our group.[44, 46]

In the present study, which is aimed at establishing the 6-membered ring TS model for the catalysis of the natural substrate by HIV-1 PR, we have used a simple model system (Scheme 3) at M06-2X[56]/6-31+G(d)[37, 56] level of theory for TS modelling and HF[36]/6-31+G(d) for quantum chemical descriptors. Using HF to estimate the energies of the frontier molecular orbitals (FMOs);[59] highest occupied molecular orbital (E_{HOMO}) and lowest unoccupied molecular orbital (E_{LUMO}), is preferred due to the delocalization error in DFT (in contrast to the tendency of HF exchange to localize charge).[60-62] These FMOs energies give rise to quantum chemical properties which are important to measure reactivity (see computational details).

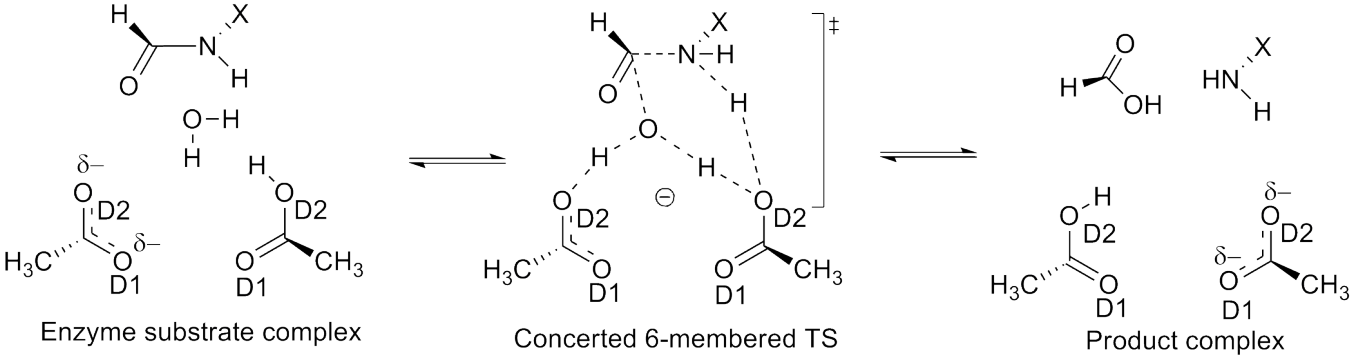
In an effort to broaden the scope of our new model, we also decided to test its applicability to a range of amide bond derivatives. Modification of the scissile bond strength was proposed based on our experience with N-methylation of amide bonds.[43, 63-65] Selective N-methylation of amide nitrogen atoms increases the proteolytic stability/bioavailability of peptides.[43, 64, 66] The methyl group is electron donating and is expected to strengthen the amide bond through a positive inductive effect, leading to an increased double bond character. On the other hand, electron withdrawing groups such as fluorine may cause the opposite effect. A better understanding of electronic methods to increase the scissile bond strength may be useful for the design of new natural substrate-based HIV-1 PR inhibitors. To achieve this in our investigation, we used

formamide and five other compounds having substituents specifically at the N-terminal plus acetic acid (to mimic protonated Asp), acetate (for unprotonated Asp) and a water molecule in the modelling of natural peptide bonds.

2 Computational details

Based on most experimental and theoretical studies, monoprotonation of one aspartate at the HIV-1 PR active site is usually adopted. For a concerted mechanism where both the nucleophile (water molecule) and electrophile (an acidic proton) are expected to attack the scissile bond synchronously, monoprotonation on the OD2 of Asp25′ is the most feasible. Hence, protonation model D (Figure 1) was considered for the study.

DFT calculations proceeded with full geometry optimization of all structures at M06-2X/6-31+G(d) level of theory in gas phase. Vibrational frequencies[67] were computed for the various unconstrained species to characterize them as local minima and transition state (TS) structures (having one-negative eigenvalue) on the potential energy surface (PES). Intrinsic reaction coordinates (IRC)[68] were computed to validate if the transition structures are truly the lowest saddle points connecting the expected structures on the reaction pathway. GaussView 5.0.8[69] was used as pre-processor and post-processor visual interface for this study and all calculations were executed within the Gaussian 09 program package.[70]



Scheme 3: Studied structures for the concerted 6-membered ring general acid-base mechanism of HIV-1 PR—substrate catalysis. X = H, F, CH_3, OCH_3, SCH_3, and $C(CH_3)_3$.

2.1 Thermochemistry and kinetics

The estimation of chemical quantities such as thermodynamic arising from energy, enthalpy and entropy contributions[71] can be achieved *via* computational modelling by normal mode analysis of the optimized structures. In this study, free energy is represented by ΔG and calculated from equation 1 as implemented in the Gaussian 09 package.[70] Its component enthalpy and

temperature-dependent changes in entropy (S) are denoted as ΔH and TΔS (at T=298.15K), respectively. Details of these quantities were discussed herein.[71, 72] Rate constant k, which is a measure of the change in concentration of the reactants or products per unit time, was obtained from equation 2 and taking the natural logarithm, was reported as lnk. Note that this parameter is only particular to the TS structure, having an activation free energy.

$$\Delta G = \Delta H - T\Delta S \qquad 1$$

To measure the effect of such electronic modifications of the C—N bond and study which substituent contributes to the strength of the scissile bond, the C—N bond strength of each TS structure was studied by estimating the adiabatic force constant. The internal force constants from Cartesian second derivatives (Hessians) were determined using M06-2X/6-31+G(d) combination and the Genadia/Matlab program[73, 74] was subsequently applied to calculate the force constant k_{CN} (equation 3).[75] The force constant is defined here as the ratio of the force required to collapse a specific covalent bond as a function of the deformation produced through the applied force.[76]

$$\mathrm{k} = \frac{\mathrm{k_B T}}{\mathrm{hc^0}} \exp\left(-\frac{\Delta \mathrm{G}^{\ddagger}}{\mathrm{RT}}\right) \qquad 2$$

Variables k, k_B, h, and R, are rate, Boltzmann, Planck, and gas constants, respectively, while c is the concentration.

$$k_{CN} = \sum_{i=1}^{3} \lambda_i^{CN} \left|\hat{u}^{CN} \times \hat{v}_i^{CN}\right| \qquad 3$$

λ_i^{CN} represents the eigenvalues of 3 × 3 interaction matrix, $\hat{u}^{CN}$ is the unit vector between atoms C and N, and $\hat{v}_i^{CN}$ are the eigenvectors of the pair interaction matrix.

2.2 Reactivity descriptors

Ab initio HF-based quantum chemical descriptors arising from the energies of the FMOs[59] which are E_{HOMO} and E_{LUMO}, respectively, were studied. These parameters were investigated to provide better insight into the nature of the TS structures, electron distribution around it, selectivity and reactivity. The properties considered are; ionization potential (IP),[77] electron affinity (EA),[77] band gap (ΔE),[77] chemical hardness (η),[78] global softness (S),[78] electronegativity (χ),[79] electrochemical potential (μ),[79] and electrophilicity index (ω)[80] (equations 3 – 8). Definitions of these quantities were extensively discussed herein.[77, 78]

$$\text{Ionization potential} \qquad \mathrm{IP} = -\mathrm{E_{HOMO}} \qquad 3$$

Electron affinity	$EA = -E_{LUMO}$	4
Band gap	$E = E_{HOMO} - E_{LUMO}$	5
Chemical hardness η	$\eta = \frac{E_{LUMO} - E_{HOMO}}{2}$	6
Global softness	$S = \frac{1}{\eta}$	7
Chemical potential	$\mu = -\frac{IP + EA}{2}$	8

3 Results and discussion

3.1 Catalysis of substrates by HIV-1 PR through the concerted 6-membered ring TS model

The proposed reaction mechanism is a concerted as well as general acid-base model. The Asp25 (acetate) is unprotonated and acts as a base while Asp25′ (acetic acid) is protonated at OD2 and acts as the acid (Scheme 3). The reaction occurs synchronously displaying the catalytic functions of the Asp groups as well as the nucleophilic role of the water molecule on the substrate, in a single step. The cleavage of the scissile bond followed the proposed reversible mechanism (Scheme 3) leading to the separated reactants/products. This process starts with the formation of an enzyme substrate (ES) complex that pre-organizes the molecules for the subsequent TS structure hence, leading to the formation of hydrolyzed products (Figure 2).

The reaction involves bond formation and breaking whereby the scissile bond is cleaved through the nucleophilic attack of water (OH) to form –COOH from the C=O of the amide bond. Two protons were also transferred within the TS structure; acetic acid gives a proton to the scissile nitrogen and water loses a proton (yielding OH) to the acetate. The overall mechanistic process remained exergonic as revealed through the negative ΔG of product complexes, PCs (Figure 2). The TS modelling strategy for the one-step concerted mechanism provides a sufficient description of the HIV-1 PR—substrate hydrolysis and the studied substrate-based derivatives.

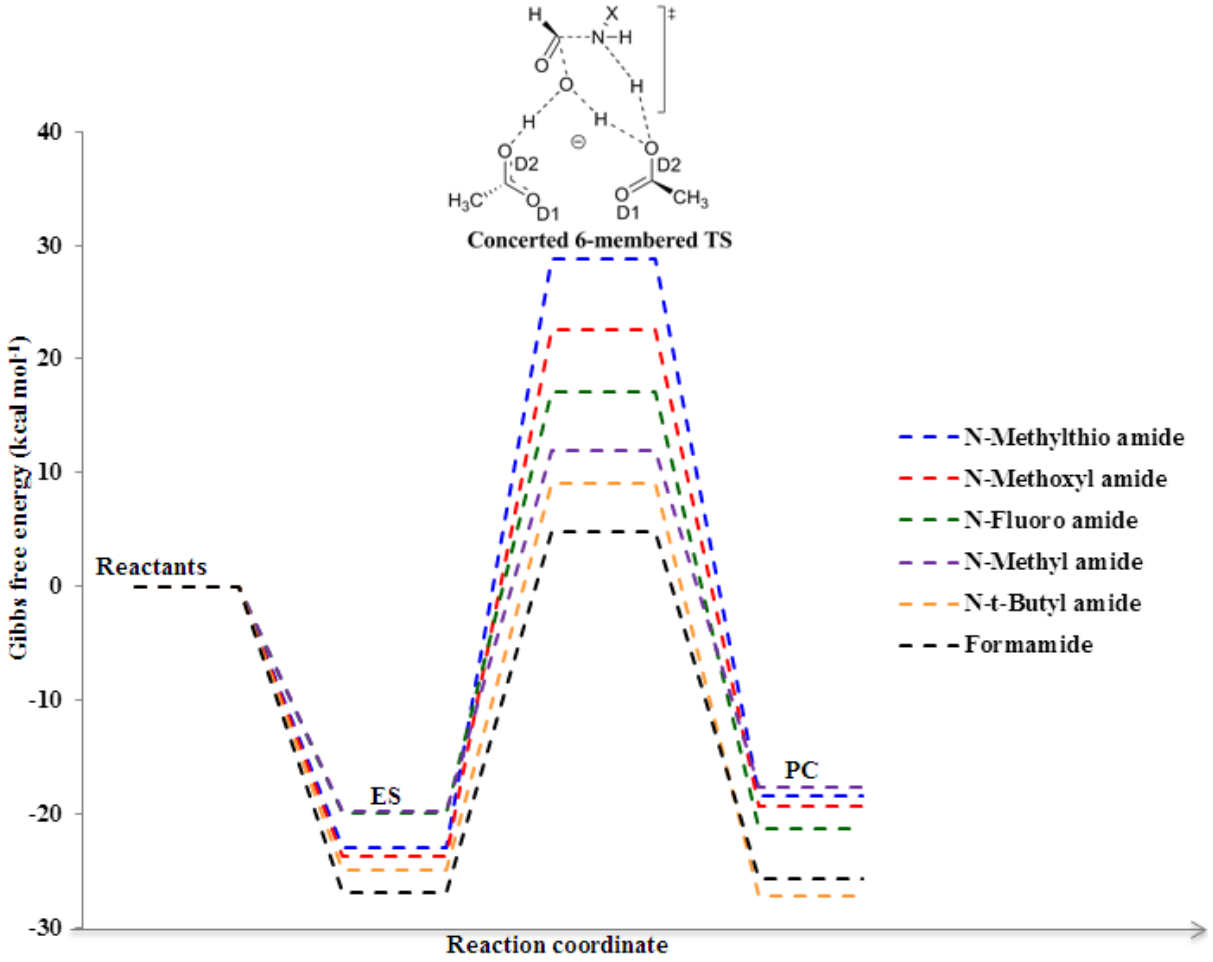


Figure 2: IRC path for the concerted cleavage of substrates in the presence of acetate, acetic acid and water using M06-2X/6-31+G(d) theoretical model. ES = enzyme substrate complex, TS = transition state structure and PC = product complex.

Presented in Table 1 are the relative thermodynamics and kinetics of the structural moieties involved in the studied reaction mechanism. All values are reported in kcal mol^{-1} relative to the sum of their separated reactants. For all the systems, the ES and PC complexes are favoured enthalpically with higher ΔH values, while the TS models are characterized by an entropy penalty. The TS structures had entropy penalties ranging from -31 to -36 kcal mol^{-1} while the enthalpy of activation ($\Delta H^{\ddagger}$) follows an increasing (more negative) trend from $SCH_3 < OCH_3 < F < CH_3 < C(CH_3)_3 < H$.

Based on accepted knowledge, the more negative the theoretical ΔG_{bind} of an enzyme—inhibitor, the greater the potential bioactivity.[81] In this study, the ES complex value is likened to the ΔG_{bind} for each system, which is within -19.8 and -26.9 kcal mol^{-1}. Also, the ΔG_{bind} of product complexes are highly negative (Figure 2 and Table 1).

The activation free energies associated with the TSs ($\Delta G^{\ddagger}$) increase (more negatively) in the same trend as $\Delta H^{\ddagger}$, giving rise to the following reaction rate order from the lnk estimation: H > $C(CH_3)_3$ > CH_3 > F > OCH_3 > SCH_3. Note that experimental $\Delta G^{\ddagger}$ was estimated to be between 15 and 21 kcal mol^{-1}[19, 10, 47-52] for peptide hydrolysis by HIV-1 PR. The calculated $\Delta G^{\ddagger}$ values for CH_3, F, OCH_3 (Table 1) fall within the experimental range while N-t-butyl amide and formamide have a much lower value than the estimated experimental data at approximately 9.0 and 5.0 kcal mol^{-1} respectively. SCH_3 has $\Delta G^{\ddagger}$ of approximately 29 kcal mol^{-1}, which is an indication of substantial slower cleavage of this substrate (SCH_3).

Table 1: Thermodynamic and kinetic parameters for the 6-membered cyclic TSs mechanism of (Scheme 3) of substrates (formamide and N-substituted derivatives) cleavage by HIV-1 PR (acetic acid and acetate) at M06-2X/6-31+G(d) theoretical model

Structure	$T\Delta S$ (kcal mol^{-1})	ΔH (kcal mol^{-1})	ΔG (kcal mol^{-1})	lnk	C—N Force constant (mDyn/A)
N-Methylthio amide					
ES	-30.09	-52.99	-22.89		
$TS^{\ddagger}$	-34.20	-5.40	28.80	-19.16	1.98
PC	-31.49	-49.91	-18.42		
N-Methoxyl amide					
ES	-30.43	-54.08	-23.65		
$TS^{\ddagger}$	-35.84	-13.20	22.64	-8.76	1.64
PC	-29.61	-48.92	-19.31		
N-Fluoro amide					
ES	-30.55	-50.43	-19.87		
$TS^{\ddagger}$	-34.24	-17.09	17.15	0.51	1.20
PC	-28.28	-45.95	-21.22		
N-Methyl amide					
ES	-29.25	-49.00	-19.77		
$TS^{\ddagger}$	-33.40	-21.51	11.89	9.39	1.13
PC	-28.28	-45.95	-17.67		
N-t-Butyl amide					
ES	-28.15	-52.99	-24.84		
$TS^{\ddagger}$	-35.13	-26.10	9.03	14.22	1.11
PC	-31.51	-58.64	-27.12		
Formamide					
ES	-28.07	-55.01	-26.94		
$TS^{\ddagger}$	-31.67	-26.86	4.80	21.36	0.85
PC	-28.02	-53.70	-25.68		

‡ denotes activation parameters $T\Delta S^{\ddagger}$, $\Delta H^{\ddagger}$ and $\Delta G^{\ddagger}$ for the TS. ES = enzyme substrate complex, TS = transition state structure and PC = product complex. Values are reported relative to the sum of separated reactants. (Cartesian coordinates of the TS structures are available in the SI).

Analysis of the electronic forces acting on the C—N bond in each TS structure provided a reliable measure of the scissile bond strength. A stronger bond usually vibrates faster than a weaker bond[76] and the adiabatic force constant provides a reliable descriptor of bond strength[82, 83]. In this investigation, the values obtained for the C—N force constant (Table 1) give a linear relation with the rate of the reaction. The lower the force constant, the faster the rate of C—N bond breakage (lower TS energy). Figure 3 gives a clear picture of the chemical reactivity of studied systems and the strength of the respective C—N bonds. There is a linear correlation between the adiabatic stretching force constant and the transition state energy (R^2=0.9) as seen in the linear fit in Figure 3. The force constant values decrease in the order; SCH_3 > OCH_3 > F > CH_3 > $C(CH_3)_3$ > H, thus showing formamide as the easiest to hydrolyse.

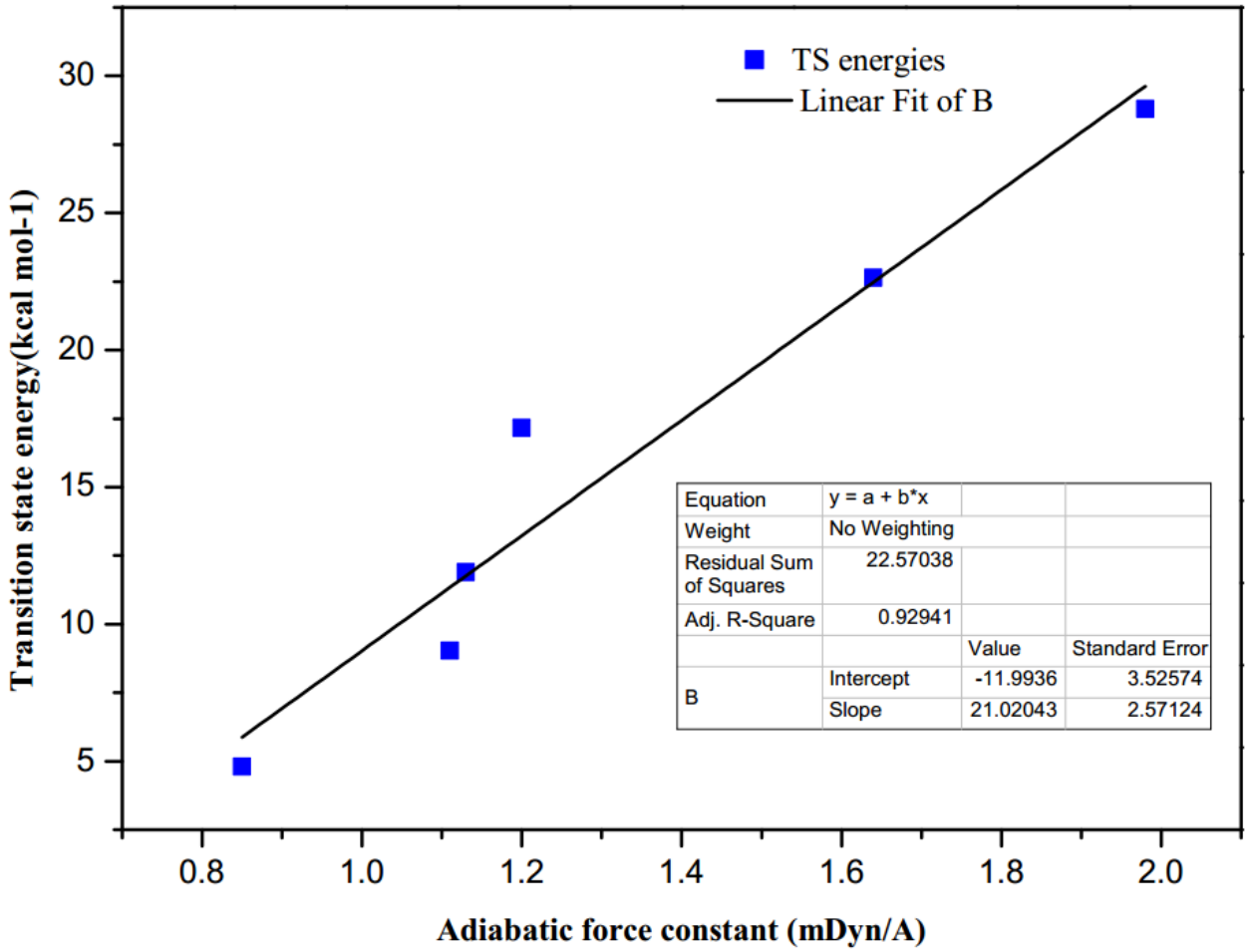


Figure 3: Adiabatic stretching force constants of C—N bonds in relation to the transition state energy calculated at M06-2X/6-31+G(d)

3.2 Quantum chemical descriptors for the concerted TS structures involved in the hydrolysis of HIV-1 PR substrates

To examine further the reactivity and selectivity of the N-substituted natural substrate derivatives of HIV-1 PR, the quantum chemical properties of each TS structures were investigated (Table 2). The energies of the HOMO and LUMO, were used to estimate the chemical reactivity, selectivity and stability of the selected derivatives using HF/6-31+G(d) level of theory. Presented in Table 2 are results of the HOMO, LUMO energies and other estimated descriptors for the mechanistic process. The lobes of the studied FMOs for selected TS structures are given in Figure 4.

Table 2: Quantum chemical properties of the TS structures involved in the one-step concerted modelling of HIV-1 PR—substrate systems at HF/6-31+G(d) level of theory

Parameter	N-Methylthio amide	N-Methoxyl amide	N-Fluoro amide	N-Methyl amide	N-t-Butyl amide	Formamide
E_{LUMO} (eV)	-0.128	-0.129	-0.130	-0.124	-0.122	-0.119
E_{HOMO} (eV)	-0.259	-0.334	-0.346	-0.358	-0.357	-0.367
IP	0.259	0.334	0.346	0.358	0.357	0.367
EA	0.128	0.129	0.130	0.124	0.122	0.119
ΔE (eV)	0.132	0.205	0.216	0.234	0.236	0.248
μ (eV)	-0.194	-0.231	-0.238	-0.241	-0.240	-0.243
η (eV)	0.066	0.102	0.108	0.117	0.118	0.124
S (eV^{-1})	15.199	9.758	9.271	8.548	8.491	8.062

IP = ionization potential, EA = electron affinity, ΔE = band gap, μ = electrochemical potential, η = chemical hardness, S = global softness.

Molecules with high E_{HOMO} values can donate their electrons easily compared to molecules with lower values,[59] therefore, the opposite holds for electron withdrawing molecules. For the present investigation, the electron withdrawing groups such as fluorine and sulphur gave lower E_{HOMO} and higher E_{LUMO} values, respectively. The overall trend for E_{HOMO} is $SCH_3 < OCH_3 < F < CH_3 < C(CH_3)_3 < H$ (same order as the rate of reaction, Table 1), which yielded IP values in this same order. The difference between these FMOs gives rise to a vital stability index called the band gap (ΔE), which measures the reactivity of the more energetically matched HOMO-LUMO pairing of the separated reactants.[59]

A small $\Delta E_{HOMO–LUMO}$ gap indicates a greater likelihood of electron transfer within the reactants involved in forming the TS structures.[78] Based on the ΔE values obtained herein, these substrates are generally stable and reactive in the order ($SCH_3 > OCH_3 > F > CH_3 > C(CH_3)_3 > H$) hence, the selected formamide and its derivatives are reactive and could be hydrolysed by HIV-1 PR. The stability of the reactive derivatives (SCH_3 and OCH_3) seems to be dependent on the

contributions of the electron donating $-CH_3$ as expected.[43, 63-65] It is quite fascinating to observe that electron-withdrawing groups (such as F) also have remarkable effect on the strength of the scissile bond compared to the electron donation N-t-butyl or formamide itself. Overall, SCH_3 has the lowest band gap, which indicates better stability leading to a slower proteolytic process; this is also reflected in its higher activation energy (Table 1).

As illustrated in Figure 4, the LUMO electrons are largely localized on the nucleophilic water molecule and the catalytic Asp groups (acetic acid and acetate). Analysis of the spatial distribution of the HOMO on each functional group of the studied substrates (Figure 4) reflects the fact that the enzymatic mechanism is indeed dependent on the selectivity of the substrate.[6] It is interesting to note that the lobes of HOMO and LUMO are the least localized in SCH_3 system compared to the other substrates.

Chemical potential (μ) and chemical hardness (η) are measures of the extent of electron delocalization and the determination of species resistance to lose electrons respectively.[77] The μ values for all of the TS structures increase from SCH_3 to H in the order; $SCH_3 < OCH_3 < F < CH_3 < C(CH_3)_3 < H$ while η follows a decreasing order of; $H > C(CH_3)_3 > CH_3 > F > OCH_3 > SCH_3$ giving rise to global softness indices (S) with values 8.1 to 15.2 eV. The chemical softness measures the reactivity of molecules.[77] The reactivity order for these TS structures decrease in the same trend as $\Delta G^{\ddagger}$; $SCH_3 > OCH_3 > F > CH_3 > C(CH_3)_3 > H$, with SCH_3 as the most reactive and formamide as the least reactive (Figure 4).

Unlike previous studies[44,46] whereby the quantum chemical descriptors correlate well with the activation free energies, it is quite fascinating to observe in this present work that an opposite trend was observe; that is, reactive derivatives have higher activation free energies. These properties could therefore, be less reliable for this model in quantifying the reactivity of the studied derivatives.

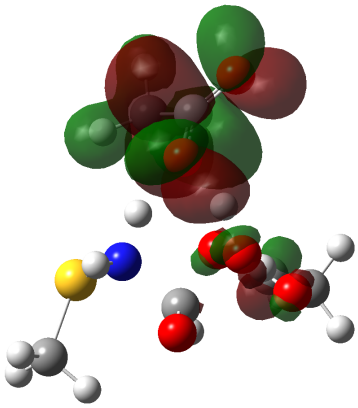
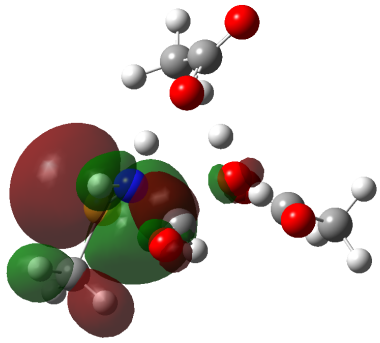
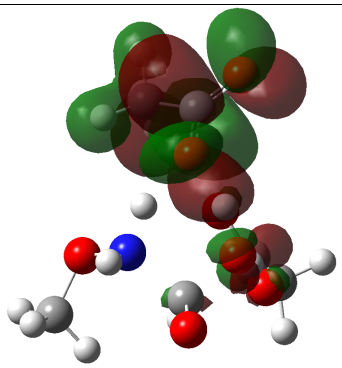
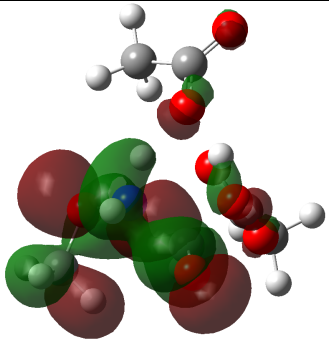
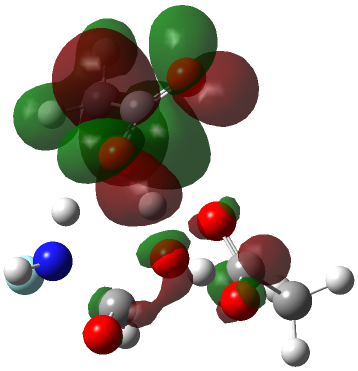
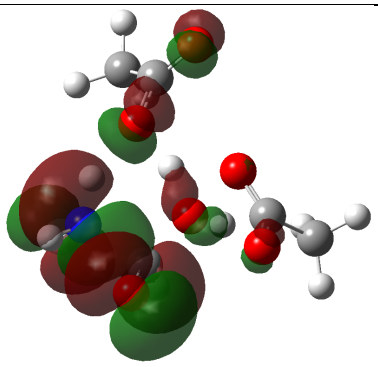
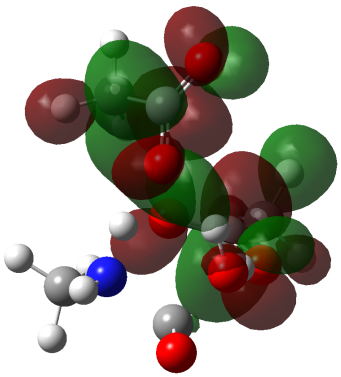
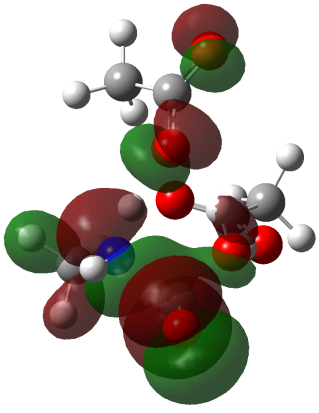
	LUMO	HOMO
N-Methylthio amide		
N-Methoxyl amide		
N-Fluoro amide		
N-Methyl amide		

Figure 4: Lobes of the frontier molecular orbitals (HOMO and LUMO) of the TS structures involved in the modelled one-step concerted hydrolysis of substrates by HIV-1 PR at HF/6-31+(G)d level of theory.

In this present work, we report a concerted mechanism that has provided a new perspective on the existence of a single rate-determining step through the 6-membered cyclic TS obtained, and activation free energies within experimental range. It is worth noting that the likelihood of the existence of more than one rate-determining step for the stepwise general acid-base mechanism of natural substrate hydrolysis by HIV-1 PRs has been suggested[21] in contrast to the most frequent opinion in which substrate separation is viewed as the rate-determining step[11, 19, 30, 84, 85] (Scheme 1).

The relevance and chemistry of sulphur as a leading constituent of pharmaceuticals was reviewed recently.[86] Amongst the organosulfur derivatives discussed are sulfonamide, sulfone, sulfinyl, thioester, thioether, thiophene, thiazole, β-lactam, thiazepine/thiazine and thiadiazole.[86] In this study, we observed N-methylthio amide (a sulfenamide) as the best substrate derivative having a negative ΔG_{bind} for its ES and reluctance to hydrolyse fast as reflected in the $\Delta G^{\ddagger}$ value (Table 1). The synthesis of such derivative with a characteristic amide-linked N—S bond formation was previously achieved using copper as a catalyst.[87, 88] Although the synthesized compounds[88] are cyclic and not HIV-1 PR natural substrate derivatives, it is quite fascinating to observe that their *in vitro* biochemical evaluation against Congenital Disorder of Glycosylation type Ia is appreciable.[88] We therefore, hope that the synthesis of natural substrate mimics with sulphur at the nitrogen of the scissile bond could possibly be a suitable inhibitor of HIV-1 PR.

4 Conclusion

The enzymatic mechanism of HIV-1 PR on natural substrate and derivatives was studied using both DFT (M06-2X) and HF methods for a small model system. The reaction mechanism was observed to follow a concerted model having one cyclic TS structure leading to product. Although a few studies have addressed the concerted general acid-base mechanism of HIV-1 PR catalysis theoretically, our investigation is the first to present the 6-membered TS model. The study has addressed mechanism through estimation of thermodynamics, kinetics, and other quantum chemical characteristics. This investigation has enabled us to deduce a correlation between the calculated activation free energies and C—N force constants of the N-substituted amide derivatives thereby, buttressing the relevance of bond strength in overcoming activation barrier. In other words, the higher the activation free energy, the higher the force constant value which also correlates to higher C—N bond strength in each substrate. The models allowed

investigation of N-substituted scissile bond derivatives, which provided us with a new potential lead (methylthio amide) for HIV-1 PR inhibition.

Within the context of this investigation, which is aimed at establishing the concerted general acid-base model for HIV-1 PR enzymatic mechanism on natural substrate and discovering a potential inhibitor through natural substrate modification, we could conclude from the title that we are not done yet on HIV-1 PR catalysis. Future studies will involve calculation of the activation energies of these compounds inside the HIV-1 PR active site.

Acknowledgements

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 Centre for High Performance Computing (www.chpc.ac.za) for computational resources.

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] 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 protease, *Journal of Molecular Biology 301*, 1207-1220.

[3] 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.

[4] 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.

[5] Ozer, N., Haliloglu, T., and Schiffer, C. A. (2006) Substrate specificity in HIV-1 protease by a biased sequence search method, *Proteins: Structure, Function, and Bioinformatics 64*, 444-456.

[6] 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.

[7] 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.

[8] 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.

[9] Hyland, L. J., Tomaszek Jr, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., and 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.

[10] 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.

[11] Rodriguez, E. J., Angeles, T. S., and Meek, T. D. (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.

[12] Co, E., Koelsch, G., Lin, Y. Z., Ido, E., Hartsuck, J. A., and Tang, J. (1994) Proteolytic Processing Mechanisms of a Miniprecursor of the Aspartic Protease of Human-Immunodeficiency-Virus Type-1, *Biochemistry 33*, 1248-1254.

[13] Schneider, J., and Kent, S. B. H. (1988) Enzymic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease, *Cell (Cambridge, Mass.) 54*, 363-368.

[14] 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.

[15] Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. J., FitzGerald, D. J., Chaudhary, V. K., Pastan, I., and Heinrikson, R. L. (1990) Interdomain hydrolysis of a truncated Pseudomonas exotoxin by the human immunodeficiency virus-1 protease, *Journal of Biological Chemistry 265*, 408-413.
[16] Louis, J. M., Nashed, N. T., Parris, K. D., Kimmel, A. R., and Jerina, D. M. (1994) Kinetics and mechanism of autoprocessing of human-immunodeficiency-virus type-1 protease from an analog of the gag-pol polyprotein, *Proceedings of the National Academy of Sciences USA 91*, 7970-7974.
[17] Kumar, M., Prashar, V., Mahale, S., and Hosur, M. V. (2005) Observation of a tetrahedral reaction intermediate in the HIV-1 protease–substrate complex, *Biochemical Journal 389*, 365-371.
[18] 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.
[19] 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.
[20] Weber, I. T., Agniswamy, J., Fu, G. X., Shen, C. H., and Harrison, R. W. (2012) Reaction Intermediates Discovered in Crystal Structures of Enzymes, In *Structural and Mechanistic Enzymology: Bringing Together Experiments and Computing* (Christov, C., and KarabenchevaChristova, T., Eds.), pp 57-86.
[21] 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.
[22] Weber, I. T., Waltman, M. J., Mustyakimov, M., Blakeley, M. P., Keen, D. A., Ghosh, A. K., Langan, P., and Kovalevsky, A. Y. (2013) Joint X-ray/neutron crystallographic study of HIV-1 protease with clinical inhibitor amprenavir: insights for drug design, *Journal of Medicinal Chemistry 56*, 5631-5635.
[23] 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*, e7860.
[24] 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.
[25] Harrison, R. W., and Weber, I. T. (1994) Molecular-dynamics simulations of hiv-1 protease with peptide substrate, *Protein Engineering 7*, 1353-1363.
[26] 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.
[27] Lee, H., Darden, T. A., and Pedersen, L. G. (1996) An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease, *Journal of the American Chemical Society 118*, 3946-3950.
[28] Liu, H. Y., MullerPlathe, F., and vanGunsteren, W. F. (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease, *Journal of Molecular Biology 261*, 454-469.

[29] Okimoto, N., Makiyama, M., Hata, M., and Tsuda, M. (1997) Gag protein hydrolysis mechanism by HIV-1 protease - Investigation by semiempirical molecular orbital method, *Journal of the Chemical Society of Japan*, 260-266.
[30] Okimoto, N., Tsukui, T., Hata, M., Hoshino, T., and Tsuda, M. (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method, *Journal of the American Chemical Society 121*, 7349-7354.
[31] Trylska, J., Bala, 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.
[32] 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.
[33] 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.
[34] Piana, S., Carloni, P., and Parrinello, M. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease, *Journal of Molecular Biology 319*, 567-583.
[35] Møller, C., and Plesset, M. S. (1934) Note on an approximation treatment for many-electron systems, *Physical Review 46*, 618.
[36] Becke, A. D. (1993) A new mixing of Hartree–Fock and local density-functional theories, *The Journal of Chemical Physics 98*, 1372-1377.
[37] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G* basis set for third-row atoms, *Journal of Computational Chemistry 22*, 976-984.
[38] Kruger, H. G. (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *Journal of Molecular Structure: Theochem 577*, 281-285.
[39] 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[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam, *Journal of Molecular Structure-Theochem 672*, 119-125.
[40] 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.
[41] 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.
[42] Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design: Molecular Modeling, Spectroscopy, and Experimental Bioactivity, *Chemical Reviews 114*, 493-537.
[43] 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.
[44] Lawal, M. M., Govender, T., Maguire, G. E., 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.
[45] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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.
[46] Lawal, M. M., Govender, T., Maguire, G. E., Kruger, H. G., and Honarparvar, B. (2017) 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.
[47] Feher, A., Weber, I. T., Bagossi, P., Boross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Torshin, I. Y., Harrison, R. W., and Tozser, J. (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites, *European Journal of Biochemistry 269*, 4114-4120.
[48] Tozser, J., Bagossi, P., Weber, I. T., Copeland, T. D., and Oroszlan, S. (1996) Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases, *Journal of Biological Chemistry 271*, 6781-6788.
[49] Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and 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*, 33231-33235.
[50] Kipp, D. R., Silva, R. G., and Schramm, V. L. (2011) Mass-Dependent Bond Vibrational Dynamics Influence Catalysis by HIV-1 Protease, *Journal of the American Chemical Society 133*, 19358-19361.
[51] Boross, P., Bagossi, P., Copeland, T. D., Oroszlan, S., Louis, J. M., and Tozser, J. (1999) Effect of substrate residues on the P2 ' preference of retroviral proteinases, *European Journal of Biochemistry 264*, 921-929.
[52] Rodriguez, E. J., Debouck, C., Deckman, I. C., Abu-Soud, H., Raushel, F. M., and Meek, T. D. (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry, *Biochemistry 32*, 3557-3563.
[53] Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E. S., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: A hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal of Biochemistry 46*, 3976-3985.
[54] 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.
[55] Becke, A. D. (1993) Densityfunctional thermochemistry. III. The role of exact exchange, *Journal of Chemical Physics 98*, 5648.
[56] Zhao, Y., and Truhlar, D. G. (2008) Density functionals with broad applicability in chemistry, *Accounts of Chemical Research 41*, 157-167.
[57] Weigend, F. (2006) Accurate Coulomb-fitting basis sets for H to Rn, *Physical Chemistry Chemical Physics 8*, 1057-1065.
[58] 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.
[59] Bradley, J. D., and Gerrans, G. C. (1973) Frontier molecular orbitals. A link between kinetics and bonding theory, *Journal of Chemical Education 50*, 463.

[60] Kronik, L., Stein, T., Refaely-Abramson, S., and Baer, R. (2012) Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals, *Journal of Chemical Theory and Computation 8*, 1515-1531.
[61] Stein, T., Eisenberg, H., Kronik, L., and Baer, R. (2010) Fundamental gaps in finite systems from eigenvalues of a generalized Kohn-Sham method, *Physical Review Letters 105*, 266802.
[62] Goerigk, L., and Grimme, S. (2010) Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions, *Journal of Chemical Theory and Computation 7*, 291-309.
[63] Zhang, S., Govender, T., Norström, T., and Arvidsson, P. I. (2005) An improved synthesis of Fmoc-N-methyl-α-amino acids, *The Journal of Organic Chemistry 70*, 6918-6920.
[64] Bose, P. P., Chatterjee, U., Hubatsch, I., Artursson, P., Govender, T., Kruger, H. G., Bergh, M., Johansson, J., and Arvidsson, P. I. (2010) In vitro ADMET and physicochemical investigations of poly-N-methylated peptides designed to inhibit Aβ aggregation, *Bioorganic & Medicinal Chemistry 18*, 5896-5902.
[65] Pratim Bose, P., Chatterjee, U., Nerelius, C., Govender, T., Norström, T., Gogoll, A., Sandegren, A., Göthelid, E., Johansson, J., and Arvidsson, P. I. (2009) Poly-N-methylated amyloid β-peptide (Aβ) C-terminal fragments reduce Aβ toxicity in vitro and in Drosophila melanogaster, *Journal of Medicinal Chemistry 52*, 8002-8009.
[66] Farahani, M. D., Honarparvar, B., Albericio, F., Maguire, G. E., Govender, T., Arvidsson, P. I., and Kruger, H. G. (2014) Proline N-oxides: modulators of the 3D conformation of linear peptides through "NO-turns", *Organic & Biomolecular Chemistry 12*, 4479-4490.
[67] Ochterski, J. W. (1999) Vibrational analysis in Gaussian, *help@ gaussian. com*.
[68] Gonzalez, C., and Schlegel, H. B. (1990) Reaction path following in mass-weighted internal coordinates, *Journal of Physical Chemistry 94*, 5523-5527.
[69] Dennington, R., Keith, T., and Millam, J. (2009) In GaussView, 5.0.8 ed., Semichem Inc., Shawnee Mission KS.
[70] 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., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Gaussian, Inc., Wallingford, CT.
[71] Ochterski, J. W. (2000) Thermochemistry in gaussian.
[72] Ryde, U. (2014) A fundamental view of enthalpy-entropy compensation, *Medchemcomm 5*, 1324-1336.
[73] Zou, W., Kalescky, R., Kraka, E., and Cremer, D. (2012) Relating normal vibrational modes to local vibrational modes with the help of an adiabatic connection scheme, *The Journal of Chemical Physics 137*, 084114.

[74] Zou, W., Kalescky, R., Kraka, E., and Cremer, D. (2013) Relating normal vibrational modes to local vibrational modes: benzene and naphthalene, *Journal of Molecular Modeling 19*, 2865-2877.
[75] Reyes, L. H., and Seminario, J. M. (2007) Determination of precise harmonic force constants for alanine polypeptides, *Journal of Molecular Structure: Theochem 818*, 125-129.
[76] Clayden, J., Greeves, N., and Warren, S. G. (2012) *Organic chemistry*, Oxford University Press, Oxford; New YorK.
[77] Zhan, C.-G., Nichols, J. A., and Dixon, D. A. (2003) Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies, *The Journal of Physical Chemistry A 107*, 4184-4195.
[78] Yang, W. T., and Parr, R. G. (1985) Hardness, Softness, and the Fukui Function in the Electronic Theory of Metals and Catalysis, *P Natl Acad Sci USA 82*, 6723-6726.
[79] Berkowitz, M. (1987) Density functional approach to frontier controlled reactions, *Journal of the American Chemical Society 109*, 4823-4825.
[80] Liu, S. B. (2005) Dynamic behavior of chemical reactivity indices in density functional theory: A Bohn-Oppenheimer quantum molecular dynamics study, *Journal of Chemical Science 117*, 477-483.
[81] Sanusi, Z. K., Govender, T., Maguire, G. E. M., Maseko, S. B., Lin, J., Kruger, H. G., 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 Modeling 76*, 77-85.
[82] Cremer, D., and Kraka, E. (2010) From molecular vibrations to bonding, chemical reactions, and reaction mechanism, *Current Organic Chemistry 14*, 1524-1560.
[83] Kraka, E., and Cremer, D. (2009) Characterization of CF Bonds with Multiple-Bond Character: Bond Lengths, Stretching Force Constants, and Bond Dissociation Energies, *ChemPhysChem 10*, 686-698.
[84] 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.
[85] 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 USA 109*, 6543-6548.
[86] Ilardi, E. A., Vitaku, E., and Njardarson, J. T. (2014) Data-Mining for Sulfur and Fluorine: An Evaluation of Pharmaceuticals To Reveal Opportunities for Drug Design and Discovery, *Journal of Medicinal Chemistry 57*, 2832-2842.
[87] Wang, Z., Kuninobu, Y., and Kanai, M. (2013) Copper-Catalyzed Intramolecular N–S Bond Formation by Oxidative Dehydrogenative Cyclization, *The Journal of Organic Chemistry 78*, 7337-7342.
[88] Dahl, R., Bravo, Y., Sharma, V., Ichikawa, M., Dhanya, R.-P., Hedrick, M., Brown, B., Rascon, J., Vicchiarelli, M., and Mangravita-Novo, A. (2011) Potent, selective, and orally available benzoisothiazolone phosphomannose isomerase inhibitors as probes for congenital disorder of glycosylation Ia, *Journal of Medicinal Chemistry 54*, 3661-3668.

Supporting Information for Chapter three

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

Monsurat M. Lawal,[a] Zainab K. Sanusi,[a] Collins U. Ibeji,[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.

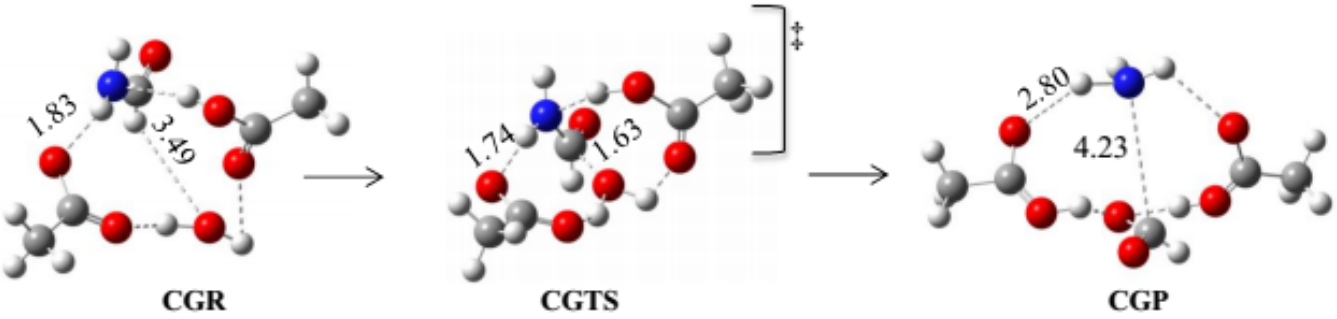
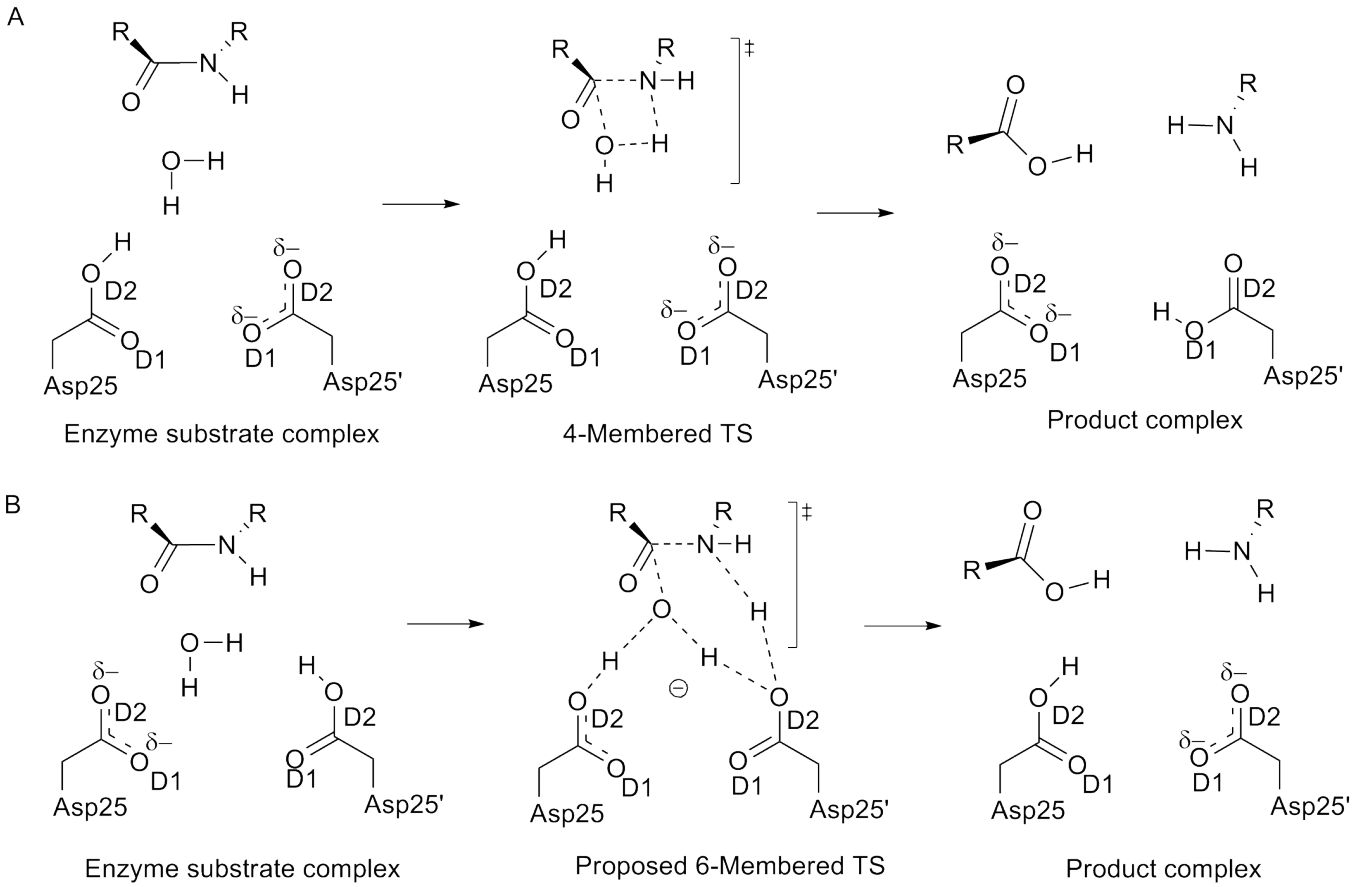


Figure S1: A pictorial illustration of the complexes involved in concerted and stepwise general acid-base mechanistic pathway for a peptide hydrolysis for a simplified HIV-1 PR active site model using MP2/6-31G(2d)/RHF/6-31G(2d).[1] The legends are: CGR = concerted general acid-base reactant, CGTS = concerted general acid-base transition state, CGP = concerted general acid-base product.



Scheme S1: (A) Concerted 4-membered ring model redrawn from literature,[2] the catalytic role of the aspartate groups seems apparently loss. (B) Proposed concerted 6-membered cyclic HIV-1 PR enzymatic mechanism.

References

[1] 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.

[2] 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.

CHAPTER FOUR

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

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.

Abstract

We give an account of a one-step concerted catalytic mechanism of HIV-1 PR hydrolysis of its natural substrate using a hybrid QM/MM method. The mechanism is a general acid-base model having both catalytic aspartate group participating and water molecule attacking the natural substrate synchronously. Three different pathways were obtained; 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 obtained 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 process.

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.

1. Introduction

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.[1-5] The aspartate dyad of the HIV-1 PR is most often monoprotonated[6] in such theoretical studies (Figure 1).

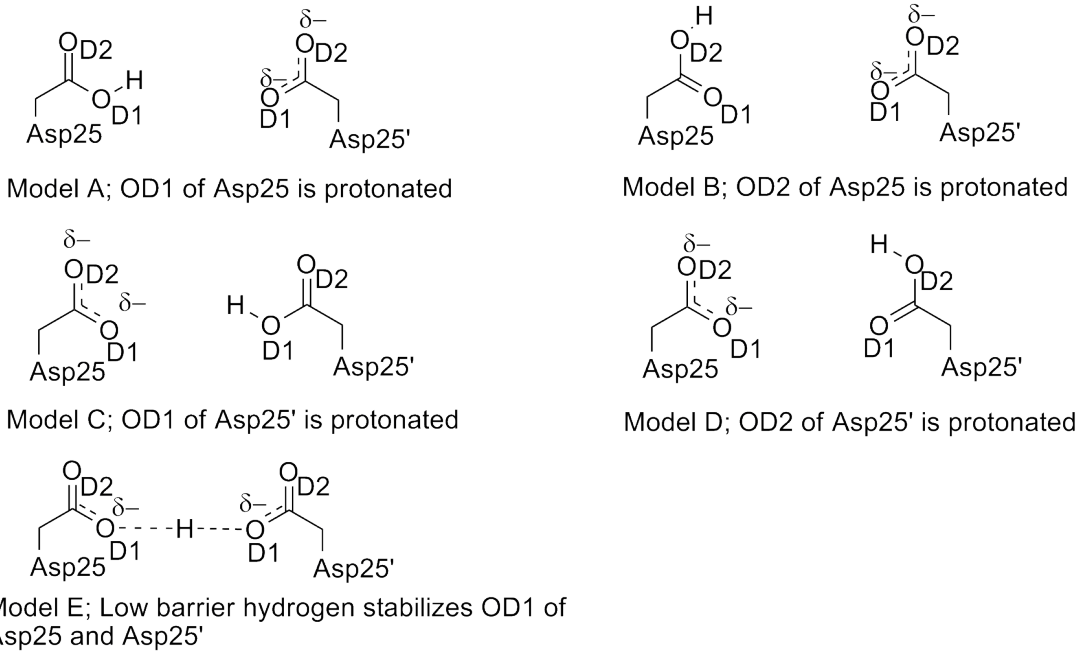
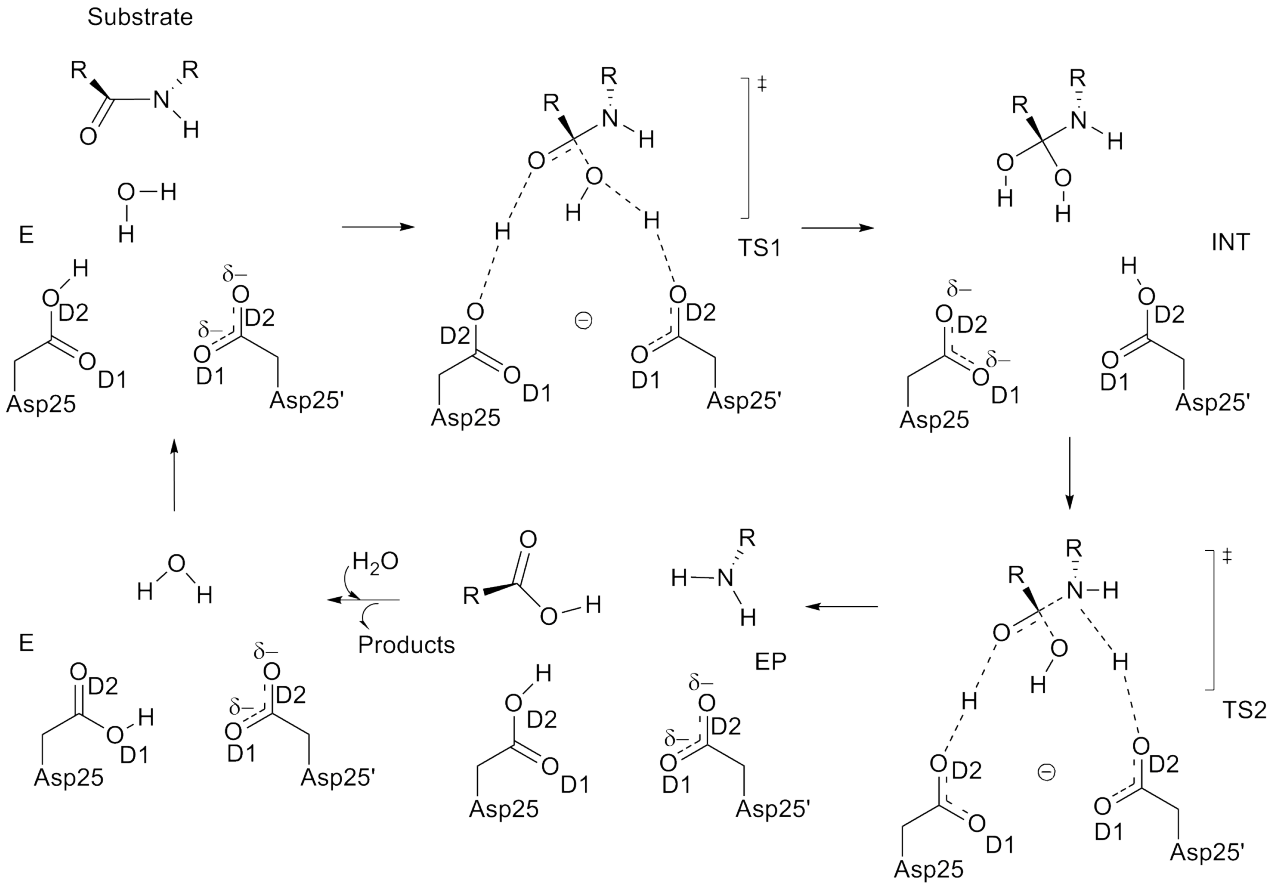


Figure 1: Possible monoprotonated models for the catalytic aspartate dyad.

Most of the recent studies[4, 7-16] 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 for the hydrolysis of the natural substrate by the PR is a nucleophilic process,[9, 10, 16] which is a variance of the stepwise general acid-base mechanism without WatC. Researchers have obtained both theoretical and experimental rate-determining transition state (TS) energies[4, 7-16] that correlate with experimental energies of peptides/substrates 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.*[3] proposed a one-step concerted mechanism for the reaction; the nucleophile (water molecule) and electrophile (an acidic proton) attack the scissile bond in a

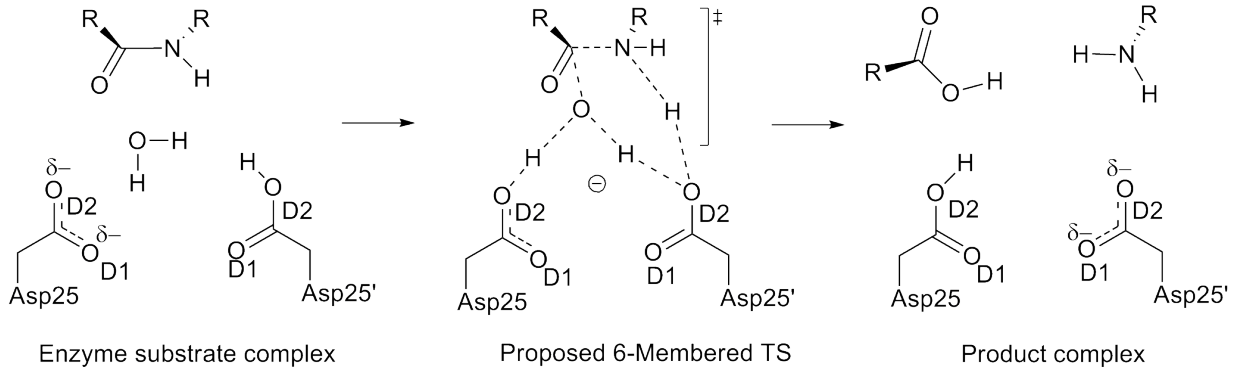
concerted manner. At the beginning of the reaction the acidic proton is located on the OD2 atom of that aspartate, which is proximal to the nitrogen atom of the scissile amide.[3] The post-reaction catalytic aspartates are still bound by the acidic proton, which now resides between the inner OD1 atoms (Model E; Figure 1). The proposed concerted mechanism in this study, was indeed derived based on the reported experimental protocol[3] which is summarized in Scheme S1 of the supporting information (SI).



Scheme 1: Stepwise general acid-base catalytic mechanism of HIV-1 PR.[17] ES is the enzyme—substrate complex, TS 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.[17]

We found only two theoretical models[4, 9] in literature with respect to the one-step concerted perspective[3] having cyclic TSs with activation free energies ($\Delta G^{\ddagger}$) of 30[9] and 43.5[4] 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.[1, 2, 18-23] In light of literature supports for theoretical 6-membered ring transition structures,[24-31] 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.



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

2. Method

2.1 System setup

The X-ray crystal structure of MA-CA substrate (code: 1KJ4)[32] was obtained from RCSB PDB[33] and complexed with HIV-1 PR (code: 1A30).[34] The catalytic water molecule was manually inserted at the active site and maintaining the distances observed from 1LV1[35] (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.[36] 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[37] (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 sites, natural substrate [Matrix-Capsid segment (MA-CA)] and water were placed at a high layer [B3LYP[38, 39]/6-31++G(d,p)[40]] while the remaining residues were at the low layer (AMBER)[41] for geometry optimization (Figure 2).

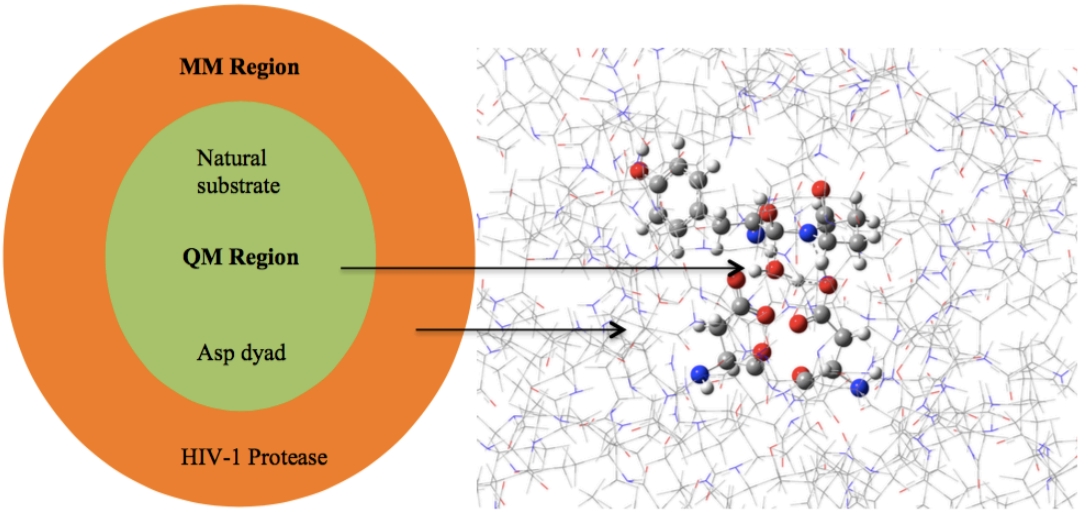


Figure 2: Schematic representation of the two-layered ONIOM (DFT:AMBER) model of HIV-1 PR—MA-CA complex. ONIOM output files for the TS structures are available in PDB format with the SI.

The reliability and accuracy of B3LYP/6-31G(d):AMBER ONIOM model has been established for a similar study.[42, 43] 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 centre, B3LYP/6-31++G(d,p):AMBER was used. Given the fact that one of the aspartate residues is protonated (Figure 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; Figure 1) as proposed from experiment.[3]

2.2 Transition state modelling and energy calculation

The complex obtained (lowest energy snapshot) from MD simulations was subjected to constraining and relaxing for the TS QM/MM modelling. 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 [B3LYP/6-31++G(d,p):AMBER] level of theory and basis set. Vibrational frequencies[44] were computed for the various species to characterize them as local minima (no negative eigenvalues) and the TS structure having exactly one imaginary frequency. Thermochemistry parameters were also obtained from the frequency calculations and the free energy, represented by ΔG was

calculated from equation 1. Intrinsic reaction coordinates (IRC)[45] 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[46] was used as pre-processor and post-processor visual interface for this study, while all calculations were executed within the Gaussian 09 program package.[47]

$$\Delta G = \Delta H - T\Delta S \qquad 1$$

Variables H, T and S are enthalpy, temperature and entropy, respectively.

3. 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 – 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 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 significantly large level of disorderliness as observed from the activation entropy ($T\Delta S^{\ddagger}$) of the TS structures having more negative values than the activation enthalpy, $\Delta H^{\ddagger}$ (Table 1). Activation free energies ($\Delta G^{\ddagger}$) of the TS structures are thus favoured by large entropy penalty 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. Sum of the total free energy of each mechanistic pathway gave -28.71, -13.87 and -46.48 kcal mol^{-1} for 6-membered cyclic, acyclic and 6-membered cyclic (with two water) systems, respectively.

Table 2: 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]

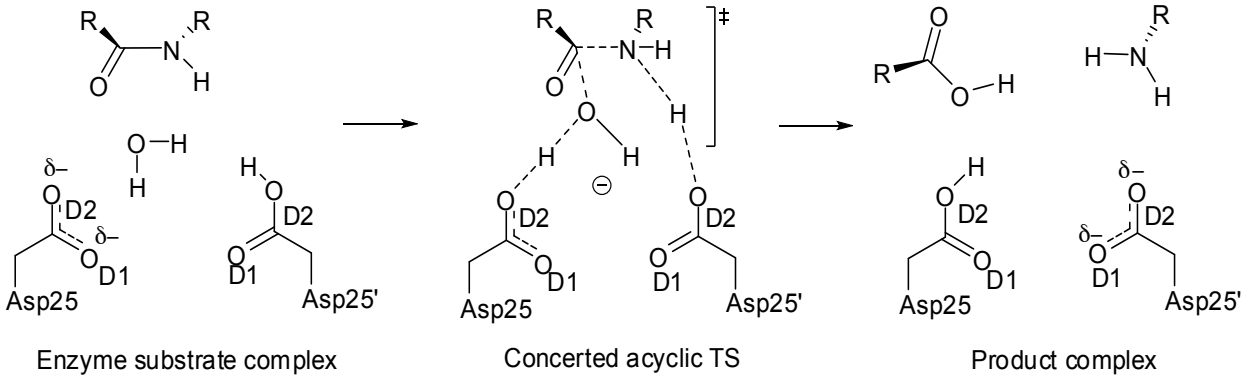
	ΔE	ΔH	TΔS	ΔG	lnk
Concerted acyclic mechanism					
ES	-40.28	-38.88	-21.66	-17.22	
C-Ac-TS	-28.60	-23.49	-38.71	15.21	3.77
PC	-34.71	-33.44	-10.74	-26.70	
Concerted 6-membered cyclic mechanism					
C-6-TS	9.45	10.23	-16.16	26.39	-15.11
PC	-43.15	-41.75	-18.71	-23.04	
Concerted 6-membered cyclic mechanism 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 kcal mol^{-1} relative to the sum of separated reactants. ES = enzyme—substrate complex, TS = transition state structure, PC = product complex, TW = two water, C = concerted and Ac = acyclic. ONIOM output files for the TS structures are available in PDB format with the SI.

Based on spontaneity (more negative ΔH) and overall system energetics (ΔE), the two-water mechanistic pathway is the most favourable; 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 $\Delta G^{\ddagger}$ of the TS structures were considered, the concerted acyclic mechanism involving a catalytic water molecule proved to be the most favoured process (Table 1). This is a well-established perspective from literature using both *in vitro* and *in silico* methods.[48] The reaction rate of the TS structures depicted by lnk, follows the order; C-Ac-TS > C-6-TS > C-6-TS-TW.

3.1 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, Figure 3) with an observed $\Delta G^{\ddagger}$ 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).



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

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.[49] The leaving proton (going 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 (Figure 4).

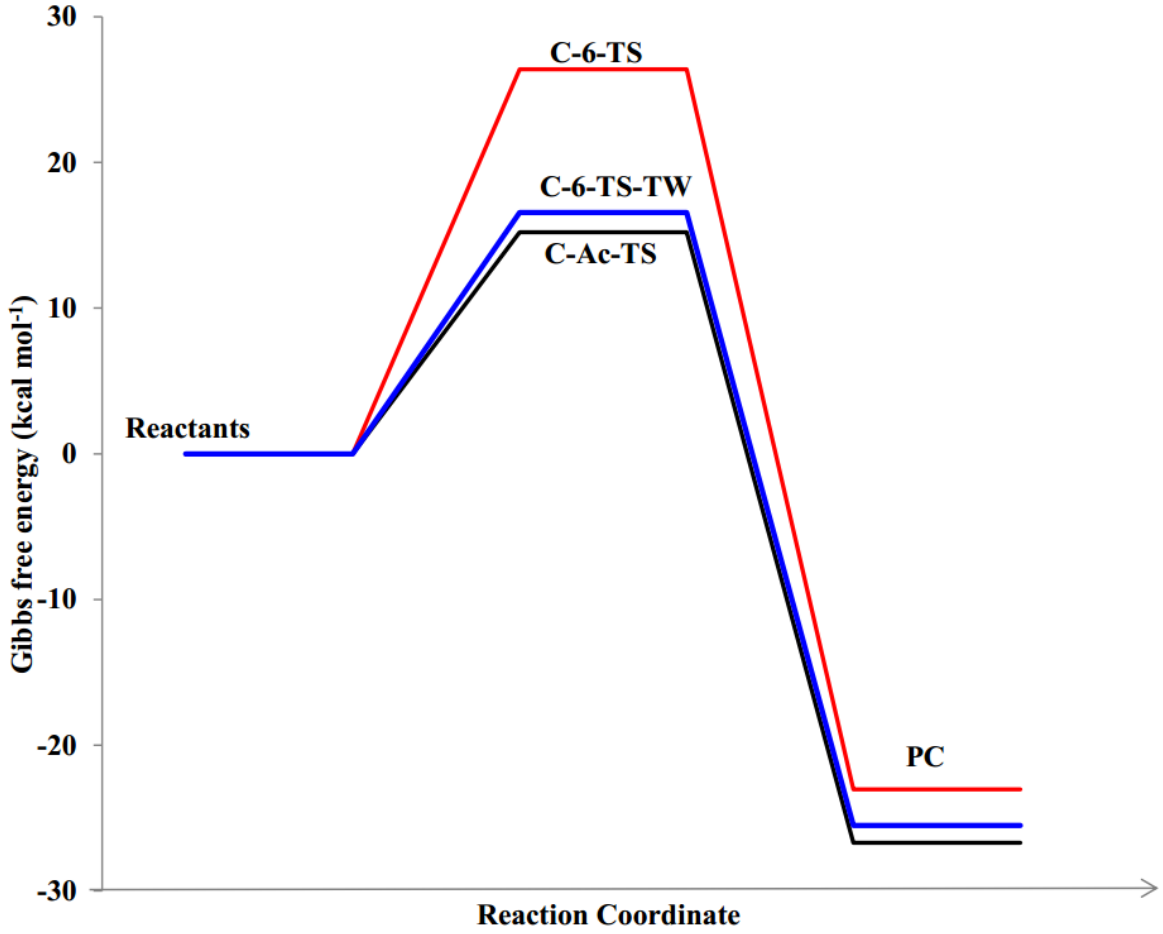


Figure 3: Free energy profile for the one-step concerted cyclic and acyclic catalytic mechanism of HIV-1 PR and MA-CA natural substrate using ONIOM [B3LYP/6-31++G(d,p):AMBER] method. TS = transition state structure, TW = two water, PC = product complex, C = concerted and Ac = acyclic. ONIOM output files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

For the acyclic mechanism, the ES gave ΔE, ΔH and TΔS 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 facilitates the concerted nature of the acyclic TS (Figure 4) with an exergonic total $\Delta E^{\ddagger}$ of -28.6 kcal mol^{-1}, $\Delta H^{\ddagger}$ value of -23.5 kcal mol^{-1}, temperature-functionalized disorderliness ($T\Delta S^{\ddagger}$) of -38.7 kcal mol^{-1} and an activation barrier ($\Delta G^{\ddagger}$) of 15.2 kcal mol^{-1}. C-Ac-TS is more favoured 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}.[20] The overall mechanism remained exergonic with the product complex having a ΔG of -26.70 kcal mol^{-1} (Figure 3 and Table 1) with a total ΔG value of -28.71 kcal mol^{-1} for the overall mechanistic process.

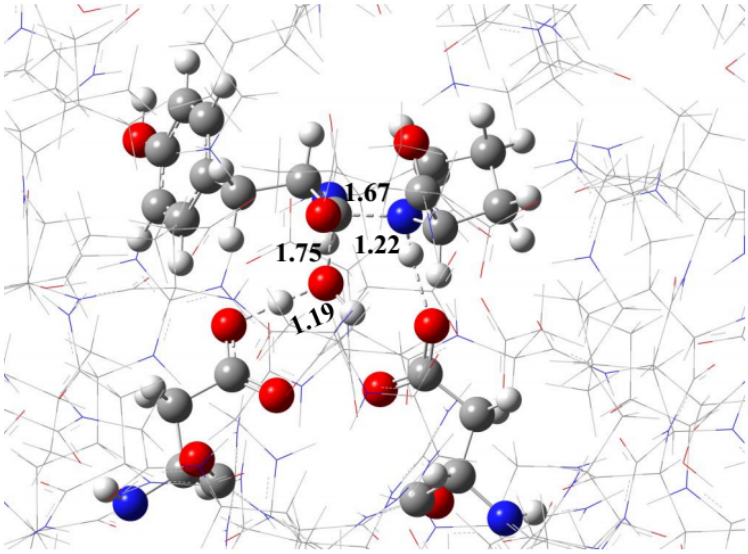


Figure 4: The concerted acyclic TS of HIV-1 PR—MA-CA general acid-base mechanism using ONIOM [DFT:AMBER] method. Picture generated with GaussView 5.0.8 [46], distances are in Angstrom (Å). ONIOM output files for the TS structures are available in PDB format with the SI (Figures S1 and S2).

3.2 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, however, 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 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′ protonation from water. In other words, only Asp25′ partakes in bond sharing and acts as 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 (Figure 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 (Figure

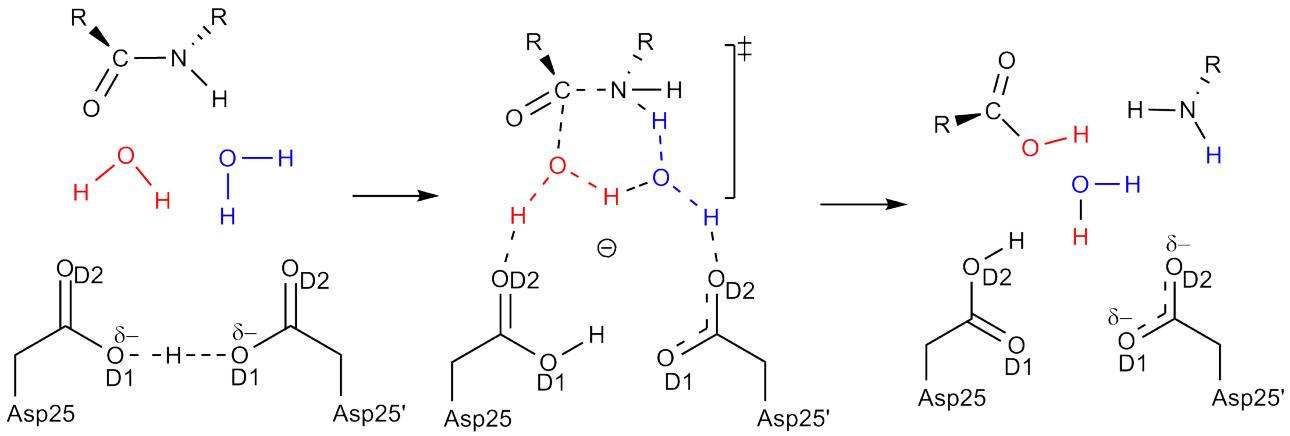
S1). The scissile nitrogen and proton from the Asp25′ gave an interatomic distance of 1.32 Å in the TS structure.

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 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 does not compare perfectly with experimental values, it is worth mentioning that this 6-membered cyclic model is an improvement on the previous contributions from literature.[4, 9] Krzemińska *et al.*[4] 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}.

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 (Figure 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.[20] This mechanism does not only showcase the success of a cyclic synchronous model but also, offers another perspective on the importance of water molecules at the active site of the HIV-1 PR[50-53] in which the activation barrier was lowered by almost 10 kcal mol^{-1} in comparison to the one water model (Figure 3).

The calculated values for ΔE, ΔH and $T\Delta S$ were -62.51, -51.66 and -22.12 kcal mol^{-1}, respectively, for ES complex of the two-water model with an estimated ΔG value was -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 TΔS, respectively, (Table 1).



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

The one-step concerted acyclic TS model provides a plausible theoretical model for the enzymatic mechanism of HIV-1 PR. Unlike previous computational attempts[4, 9] 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 favourable 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;[54] MA-CA scissile bond is located between Tyr-Pro (Figure 2).

4. 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.[55]

These models also provide new 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

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

Conflicts of interest

The authors declare no competing interests.

Supporting information

Proposed concerted scheme based on experimental report, 3D pictorial representation of the 6-membered cyclic TSs and the PDB formats of the ONIOM output files for the TS structures are available here.

Notes and Reference

[1] Hyland, L. J., Tomaszek Jr, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., and 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.

[2] 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.

[3] 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.

[4] 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.

[5] Ribeiro, A. J. M., Santos-Martins, D., Russo, N., Rarnos, M. J., and Fernandes, P. A. (2015) Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease, *ACS Catalysis 5*, 5617-5626.

[6] Mcgee, T. D., Edwards, J., and Roitberg, A. E. (2014) pH-REMD Simulations Indicate That the Catalytic Aspartates of HIV-1 Protease Exist Primarily in a Monoprotonated State, *Journal Physical Chemistry B 118*, 12577-12585.

[7] Rodriguez, E. J., Angeles, T. S., and Meek, T. D. (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.

[8] 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.

[9] 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.

[10] 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.

[11] Lee, H., Darden, T. A., and Pedersen, L. G. (1996) An ab initio quantum mechanical model for the catalytic mechanism of HIV-1 protease, *Journal of the American Chemical Society 118*, 3946-3950.

[12] Liu, H. Y., MullerPlathe, F., and vanGunsteren, W. F. (1996) A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease, *Journal of Molecular Biology 261*, 454-469.

[13] Okimoto, N., Makiyama, M., Hata, M., and Tsuda, M. (1997) Gag protein hydrolysis mechanism by HIV-1 protease - Investigation by semiempirical molecular orbital method, *Journal of The Chemical Society of Japan*, 260-266.

[14] Okimoto, N., Tsukui, T., Hata, M., Hoshino, T., and Tsuda, M. (1999) Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method, *Journal of the American Chemical Society 121*, 7349-7354.

[15] Trylska, J., Bala, 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.
[16] 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.
[17] Piana, S., Carloni, P., and Parrinello, M. (2002) Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease, *Journal of Molecular Biology 319*, 567-583.
[18] Feher, A., Weber, I. T., Bagossi, P., Boross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Torshin, I. Y., Harrison, R. W., and Tozser, J. (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites, *European Journal of Biochemistry 269*, 4114-4120.
[19] Tozser, J., Bagossi, P., Weber, I. T., Copeland, T. D., and Oroszlan, S. (1996) Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases, *Journal of Biological Chemistry 271*, 6781-6788.
[20] Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and 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*, 33231-33235.
[21] Kipp, D. R., Silva, R. G., and Schramm, V. L. (2011) Mass-Dependent Bond Vibrational Dynamics Influence Catalysis by HIV-1 Protease, *Journal of the American Chemical Society 133*, 19358-19361.
[22] Boross, P., Bagossi, P., Copeland, T. D., Oroszlan, S., Louis, J. M., and Tozser, J. (1999) Effect of substrate residues on the P2 ' preference of retroviral proteinases, *European Journal of Biochemistry 264*, 921-929.
[23] Rodriguez, E. J., Debouck, C., Deckman, I. C., Abu-Soud, H., Raushel, F. M., and Meek, T. D. (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry, *Biochemistry 32*, 3557-3563.
[24] Kruger, H. G. (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *Journal of Molecular Structure: Theochem 577*, 281-285.
[25] 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[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam, *Journal of Molecular Structure -Theochem 672*, 119-125.
[26] 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.
[27] Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E. S., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: A hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal of Medicinal Chemistry 46*, 3976-3985.
[28] 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.
[29] 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.
[30] Lawal, M. M., Govender, T., Maguire, G. E., 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.
[31] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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.
[32] 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.
[33] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank, *Nucleic Acids Research 28*, 235-242.
[34] Louis, J. M., Dyda, F., Nashed, N. T., Kimmel, A. R., and Davies, D. R. (1998) Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease, *Biochemistry 37*, 2105-2110.
[35] 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.
[36] Fakhar, Z., Govender, T., Maguire, G. E., Lamichhane, G., Walker, R. C., Kruger, H. G., and Honarparvar, B. (2017) Differential flap dynamics in L, D-transpeptidase2 from Mycobacterium tuberculosis revealed by molecular dynamics, *Molecular BioSystems 13*, 1223-1234.
[37] 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.
[38] 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.
[39] Becke, A. D. (1993) A new mixing of Hartree–Fock and local density-functional theories, *The Journal of Chemical Physics 98*, 1372-1377.
[40] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G* basis set for third-row atoms, *Journal of Computational Chemistry 22*, 976-984.
[41] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.
[42] Sanusi, Z. K., Govender, T., Maguire, G. E., 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.
[43] Sanusi, Z., Govender, T., Maguire, G., 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.
[44] Ochterski, J. W. (1999) Vibrational analysis in Gaussian, *help@ gaussian. com*.

[45] Gonzalez, C., and Schlegel, H. B. (1990) Reaction path following in mass-weighted internal coordinates, *Journal of Physical Chemistry 94*, 5523-5527.
[46] Dennington, R., Keith, T., and Millam, J. (2009) In GaussView, 5.0.8 ed., Semichem Inc., Shawnee Mission KS.
[47] 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., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Gaussian, Inc., Wallingford, CT, USA.
[48] Lawal, M. M., Sanusi, Z. K., Govender, T., Maguire, G. E. M., Honarparvar, B., and Kruger, H. G. (2018) From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets, *Current Medicinal Chemistry*.
[49] 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.
[50] Liu, Z., Wang, Y., Yedidi, R. S., Dewdney, T. G., Reiter, S. J., Brunzelle, J. S., Kovari, I. A., and Kovari, L. C. (2013) Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes, *Biochemical and Biophysical Research Communications 430*, 1022-1027.
[51] Li, Z., and Lazaridis, T. (2003) Thermodynamic contributions of the ordered water molecule in HIV-1 protease, *Journal of the American Chemical Society 125*, 6636-6637.
[52] Wang, Y.-X., Freedberg, D. I., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., Bhat, T. N., Erickson, J. W., and Torchia, D. A. (1996) Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR, *Journal of the American Chemical Society 118*, 12287-12290.
[53] 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.
[54] Tozser, J., Zahuczky, G., Bagossi, P., Louis, J. M., Copeland, T. D., Oroszlan, S., Harrison, R. W., and Weber, I. T. (2000) Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases, *European Journal of Biochemistry 267*, 6287-6295.
[55] 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.

Supporting Information for Chapter Four

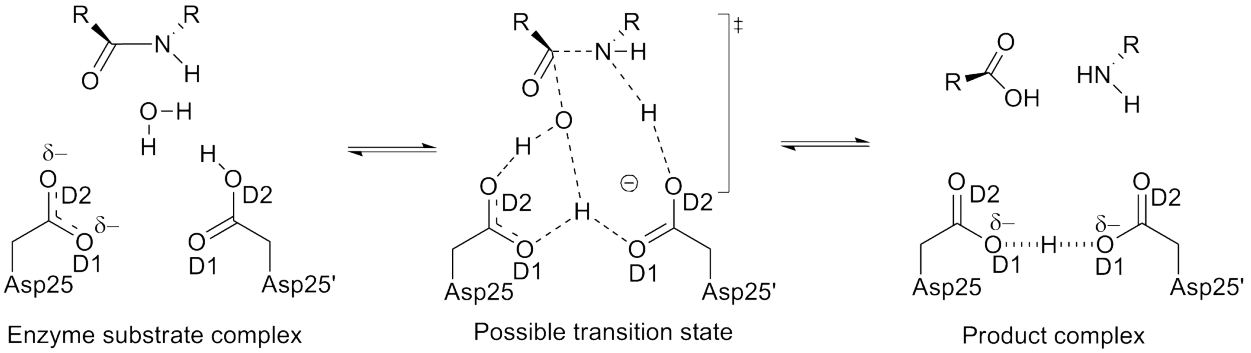
Unravelling the concerted catalytic mechanism of the Human Immunodeficiency Virus type 1 (HIV-1) protease: a hybrid QM/MM study

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.

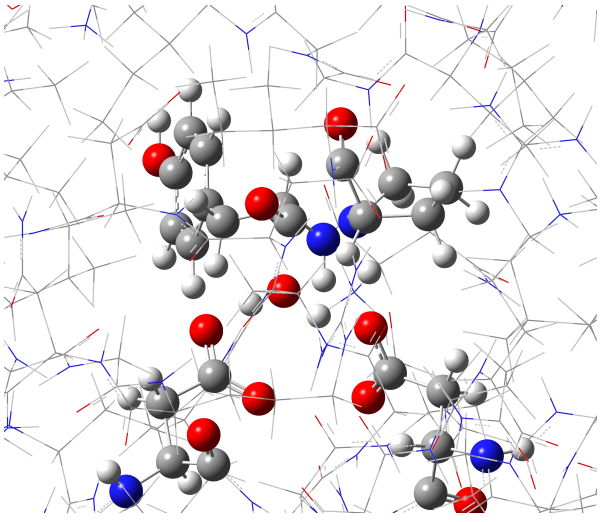


Figure S1: Concerted 6-membered cyclic TS involving one water molecule at the active site of HIV-1 PR—MA-CA complex

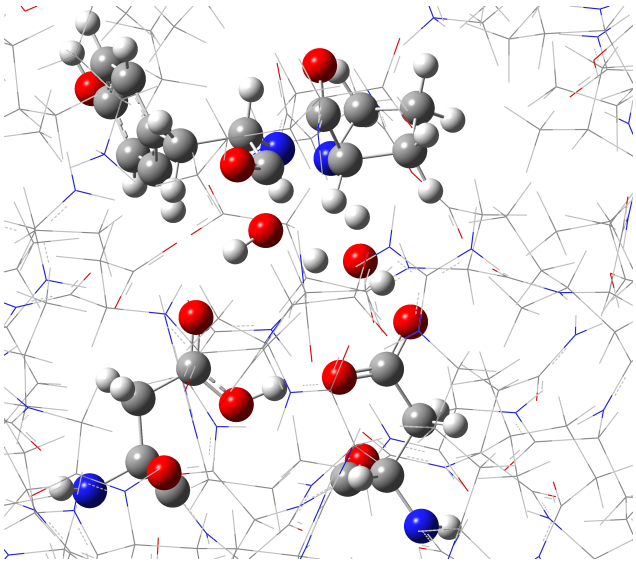


Figure S2: Concerted 6-membered cyclic TS involving two water molecules at the active site of HIV-1 PR—MA-CA complex

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.

CHAPTER FIVE

A Computational 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.

Abstract

The Human Immunodeficiency Virus type 1 (HIV-1) protease is a crucial target for HIV/AIDS treatment and our understanding of its catalytic mechanism is the basis on which HIV-1 enzyme inhibitors are developed. Although, several experimental studies indicate that HIV-1 protease facilitates the cleavage of the Gag and Gag-Pol polyproteins and it is highly selective with regards to the cleaved amino acid precursors, physical parameters, the main principles of substrate specificity and recognition, remain poorly understood. By means of transition state modelling, using a one-step concerted mechanistic process, 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 non-scissile) within the polypeptide sequence using ONIOM calculations. It was observed that the studied HIV-1 PR subtypes recognize and cleave at both scissile bond and non-scissile domains with characteristic lower activation barriers in the default scissile bond regions. In addition, peptide sequence is an important factor in the substrate recognition and specificity for HIV-1 PR, a model substrate sequence of at least ten amino acid residues is required for suitable recognition and cleavage.

Keywords: HIV-1 subtypes B and C-SA PR; cleavage mechanism; Our Own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM); natural substrates; activation free energy, concerted transition state.

1 Introduction

A group of proteolytic enzymes with two aspartyl (Asp) residues at their active sites are known as the aspartic proteinases and these include; renin,[1, 2] pepsin,[1, 2] penicillopepsin[3] and the human immunodeficiency virus protease (HIV-1 PR).[3, 4] HIV-1 PR catalyses the hydrolysis of Gag and Gag-Pol polyprotein precursors at specific sites (scissile bonds), which fulfil an essential role in the maturation of infectious virus that causes AIDS in humans.[5-7] 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.[8-10] After the 3-dimensional structure of the HIV-1 PR was resolved using single X-ray analysis,[11, 12] numerous studies have been performed on the enzyme over the last few decades, to design specific and potent inhibitors.[13-15] Despite that, the detailed mechanism of HIV-1 PR catalysis of the natural substrate is still a matter of controversy.[16] One factor that has impaired a comprehensive analysis is the different sequence homology of the cleavage sites, making the substrate specificity determinants difficult to identify.[17] In addition, the cleavage sites sequence that occur naturally have been compared for HIV-1 versus HIV-2 and results reveal that both enzymes do not have a precise consensus substrate sequence.[18, 19] Since the transition state (mechanism-based) analogue of the enzyme-catalysed reaction is a significant type of inhibitor, it is crucial to fully understand the mechanism of the HIV-1 enzyme.[20]

Several proposals for the HIV-1 PR cleavage mechanism based on both *in vitro* and *in silico* work have been suggested,[3, 4, 13, 21-24] however, there is no consensus yet on the most feasible mechanistic pathway. Most of these studies classify the mechanism into two general groups; the nucleophilic process[4, 10, 22] and the general acid-general base,[13, 25, 26] the latter is the predominant mechanism from literature for aspartic proteases and HIV-PR in particular.[16, 24, 27-30] The nucleophilic mechanism involves a direct attack on the scissile peptide (C—N) bond by one of the protonated catalytic Asp that acts as the nucleophile,[7] yielding an anhydride intermediate (stepwise nucleophilic reaction) or a transition state leading to product (concerted nucleophilic reaction).[4, 20] In the general acid-base process, a catalytic water molecule acts as the nucleophile as earlier proposed by Fruton[26] and Davies *et al*.[31] 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 Asp protease, this proceeds with the protonated Asp donating its hydrogen to the nitrogen of the scissile bond.[32] The unprotonated Asp ionizes the catalytic water by accepting its

proton, and this process can either be a stepwise or a concerted reaction in which the former has been considered the most likely from literature.[4, 10, 33] The HIV-1 PR as well as their active mutants recognizes and cleaves the same polypeptides precursor, therefore, there remains a mutual structure-function relationship among them that has been, and should be exploited in existing and future drug-inhibitors.[34]

Another major debate on the HIV-1 PR reaction mechanism is the protonation state of the catalytic Asp and this has been discussed extensively in literature.[16, 35-37] Since both experimental and theoretical studies show that the HIV-1 catalytic process function at a pH of 2-7[36, 38, 39], hence, the protonation state/models proposed in our review[16] is plausible. In the context of this study, PROPKA[40] was used to assign the protonation state for the catalytic Asp and a mono-protonation model was selected as supported by most studies.[4, 13, 20, 24, 39, 41]

Earlier theoretical studies that have focused on the catalytic mechanism of the HIV-1 PR, utilized molecular mechanics (MM) and molecular dynamics (MD) modelling to study the enzyme—substrate and enzyme—intermediate complexes.[42-44] However, one of these studies used a model system that focused mainly on the active site and the substrates,[20] while the other only examined the active site with its closest surrounding groups.[42, 45] Another limitation was the use of modified substrate sequences similar to the natural substrate, in some cases six-alanine residues were used for the substrate sequence, and in other cases one or more of the amino acids was substituted for another amino acids.[10, 34, 46]

Density functional theory (DFT) methods have been considered suitable to elucidate mechanistic pathways and obtain results closer to experimentally derived values.[35, 47-49] The reactivity of a HIV-1 PR model was examined by Garrec *et al.* in 2011, specifically to test the accuracy of the latest DFT methods.[35] They suggested the combination of B3LYP[50]/6-31++G(d,p)[51, 52] or B3LYP/6-311++G(d,p) for geometry optimization, and the former theoretical level was adopted herein.

It is worth noting that two theoretical approaches[4, 46] with regards to the one-step concerted mechanism for HIV PR cleavage of natural substrates are found in literature. In these studies cyclic transition states were employed, yielding activation energies ($\Delta G^{\ddagger}$) of 30.0[4] and 43.5[46] kcal mol^{-1}. The experimental $\Delta G^{\ddagger}$ values, depending on the peptide sequence, range between 14.86 and 21.03 kcal mol^{-1}.[13, 19, 53-57]

In this study, we aimed to continue the investigation of the one-step concerted reaction mechanism of the HIV-1 PR from our previous work.[58] Two enzymes, HIV-1 subtype B[59] and a popular mutant in South Africa (subtype C-SA)[60] complexed with matrix-capsid, MA-CA (Gag) and RNAseH-integrase, RH-IN (Gag-Pol) natural substrates were used herein. Likewise, the Gag and Gag-Pol polyprotein sites that are not normally cleaved (non-scissile domains) were also considered. This study was designed to obtain a better understanding of substrate recognition by estimating the activation free energies at different peptide bond regions in addition to the scissile bond domain within the same substrate sequence. We also investigated the substrate length of 6 amino acid residues versus 10 peptide sequences for a more accurate recognition. The recognition sequences cleaved by HIV-1 PR are represented in the supporting information (Table S1).

The catalytic process follows a concerted and general acid-base mechanism, with mono-protonation, same as the model used in our previous work.[24, 58, 61, 62] Since the hydrolysis by HIV-1 PR involves the breaking and forming of chemical bonds, classical molecular mechanics is insufficient for such an investigation. A combined quantum mechanics/molecular mechanics (QM/MM) potential was rather applied.[7, 10, 63-65] We therefore, report a study using Our Own N-layered Integrated molecular Orbital and molecular Mechanics (ONIOM),[66-68] a hybrid QM/MM algorithm.

2 Method

2.1 The system set-up

Initial structure coordinates of the subtype B and C-SA HIV-1 PR were taken at 1.8 Å and 2.7 Å resolution from the Brookhaven Protein Data Bank (PDB codes:1HXW[59] and 3U71[60]), respectively, while the structure coordinates for the substrates considered (MA-CA[59] and RH-IN[60]) is a P3-P3' amino-acid sequence, PDB codes:1KJ4 and 1KJ, respectively.[69] The MA-CA crystal complex was obtained at a 2.9 Å resolution,[69] while the remaining crystal complexes used (RH-IN and catalytic water) had a 2.0 Å resolution.[69, 70]

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 selected non-scissile domains of the same natural substrate). A major discrepancy was discovered at the P2 position for RH-IN, in some cases valine was used by some groups[71, 72] while others used Ile.[73-76] In this study, Ile was used because this is the most common amino acid

sequence used[73-76] and the 3D structure is available in PDB bank.[73] 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, while -Phe*Leu- and -Arg*Lys- are non-scissile 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, with the cleavage domain in the centre (Table S1), afterwards, we proceed to a P5-P5' ten peptide sequences because of some discrepancy found for RH-IN.

As water is an important factor for the HIV-1 PR proteolysis, a catalytic water molecule closest to the active site is taken from the PDB code:1LV1[70] and superimposed on both enzyme—substrate complexes considered, this is represented in Figure 1. 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 complexes considered for consistency. The position of the substrate with the catalytic water in the active site binds similar to the HIV PR—inhibitor complexes[61, 62] and this was superimposed using PyMOL.[77] Thus, the crystallographic waters and ions present within the structures of the entire enzyme—substrate complexes were removed leaving only the catalytic water molecule at the active site.

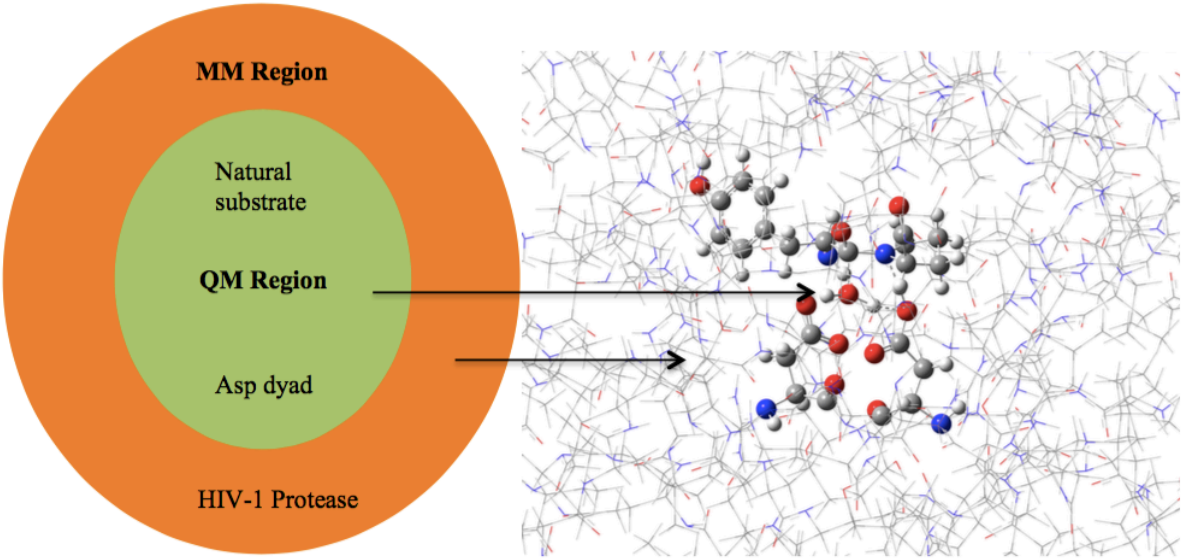


Figure 1: Schematic representation of the applied two-layered ONIOM model (B3LYP/6-31++G(d,p):AMBER) for HIV-1 PR—MA-CA (P3-P3′) complex (acyclic TS with one water molecule). All the 3D structures of the enzyme—substrate complexes considered are provided with the supporting information (Gaussian input file format).

The pKa values of the titratable amino acids were calculated using the empirical PROPKA,[40, 77] and a mono-protonated Asp25/25' state was induced in the binding site at physiological pH 7 using GaussView.[78] Note that the standard pKa values of ionizable groups can be shifted by protein environments.[79] Thereafter, the enzyme—substrate complexes considered with the Asp proton (H^+) positioned midway to the nitrogen of the scissile bond and this position has been shown to be stable energetically via molecular dynamics (MD) simulations.[44]

The first results (Table 1) for the P3-P3' RH-IN substrate revealed lower energies for some of the non-scissile bonds. This prompted us to also study a longer substrate (P5-P5') as well as performing a constrained MD simulation on the substrate to potentially obtain better recognition between the substrates and the enzyme prior to the transition state calculation.

The MD simulation protocols for P5-P5' HIV PR complexes proceed as described by Ribeiro *et al*.[80] This involves minimizations at 2500 cycles, and at the heating stage constraint was applied to counteract the SHAKE failure that might result as an increase in temperature from 0K to 600K (this is necessary to explore all possible enzyme—substrate conformations). A stepwise equilibration was done in 3 stages; 10, 5 and 2, and the constraint forces was reduced at each step. In the production stage, 3-4 Å constraint distance between the OH of the water and carbonyl of the scissile bond and the OH of the protonated Asp with the nitrogen of the scissile bond was fixed and run for 10 ns. Thereafter, an unconstrained MD calculation for 2 ns was done, and a representative snapshot of the lowest energy conformation was obtained.

The lowest energy conformations for all complexes were superimposed with the enzymes (B and C-SA, respectively) that did not undergo constraint MD, and was taken for further ONIOM calculations in which the acyclic concerted TS starting structures[58] were constructed using modred; interatomic distances were constrained (Owat--Csub, Nsub--OD2 Asp25′ and Owat--OD2 Asp25′) and optimised to construct a suitable starting structure for an unconstrained transition state optimization.

The active Asp25/25' residues, catalytic water, as well as the substrate were considered at a high layer (QM/B3LYP[50, 81]/6-31++G(d,p)[51, 52]), while the remaining system was considered at a lower (MM-AMBER[82]) layer for the ONIOM calculations. Since the reaction centre involves the transfer of proton due to transition state modelling, the electric field is taking into consideration thereby specifying an ONIOM electronic embedding conformation.[83]

2.2 Free energy calculations

To model the breaking and forming of chemical bonds, electrons are included explicitly in the energy calculations,[71] and to achieve this in an entire enzyme, a hybrid method was employed.[7] Hence, the enzyme was divided into two regions, one classical and quantum mechanical,[20, 63, 65] it is important for the later to describe the chemistry of the reaction process correctly.[15] A two-layer ONIOM[68, 84] approach was thus applied to calculate the activation free energy of the catalysed reaction. Preceding studies showed hybrid QM/MM methods are adequate for HIV-1 PR catalytic mechanism[7, 63-65] and the accuracy of DFT/B3LYP method has been tested for HIV-1 PR reactivity.[46] All complex structures for obtaining starting optimized reactant states are prepared from energy minimization calculation using the Amber forcefield,[85] 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 gas phase at a B3LYP/6-31++G(d,p):AMBER level of theory, vibrational frequencies[86] for the different unconstrained complexes were computed to characterize them as TS structures with one negative eigenvalue. In this study, GaussView 5.0.8[78] was used for both pre and post-processor graphic interface analyses, while all calculations were performed within the Gaussian 09 package.[87]

Estimating the chemical quantities such as thermodynamics that arises from energy, entropy contribution and enthalpy could be achieved computationally.[88] In the context of this study, the activation free energy, enthalpy and temperature-dependent (at T=298.15K) change in entropy is denoted as: $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$ and $T\Delta S^{\ddagger}$, respectively. Provided in equations 1–3 are the activation free energy attained from frequency calculations for the studied polypeptide sequences involving the scissile bond and non-scissile bond, as well as their difference within the enzyme—substrate system of both subtypes B and C-SA HIV-1 PR, respectively.

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

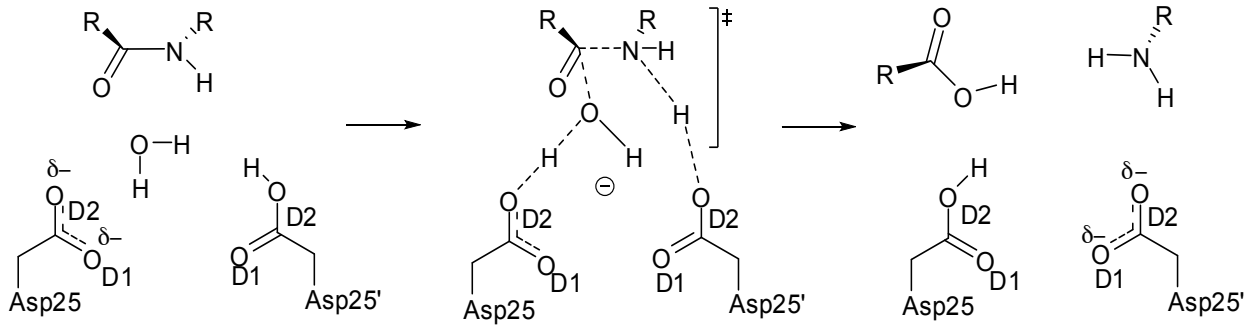
$$\Delta G^{\ddagger}_{non-scissile} = \Delta G_{complex} - \Delta G_{protease} - \Delta G_{non-scissile} \quad (2)$$

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

3 Results and discussion

3.1 Recognition and catalysis of substrates by HIV-1 PRs

By means of TS modelling, the 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 1). The catalytic function of the aspartate groups combined with the nucleophilic attack of the water molecule on the C—N scissile bond and other amide bonds occurred concurrently. Thus, the concerted acyclic process starts with the catalytic water (OH) attacking the carbonyl atom of the scissile bond forming COOH. The Asp25' provides a proton to the nitrogen atom of the scissile bond becoming a base, while Asp25 accepts proton to yield OH from the nucleophilic water and becomes acidic. The cleavage of the scissile (or non-scissile) peptide bond is a reversible mechanism as earlier proposed,[58] which led to the separated reactants and products. We have previously, reported the acyclic concerted theoretical mechanism,[58] which corresponds excellent with reported experimental results.



Scheme 9: Concerted acyclic enzymatic mechanism of HIV-1 PR with its substrates.

The overall mechanistic process is a spontaneous reaction as observed from $\Delta E^{\ddagger}$, $\Delta H^{\ddagger}$ and $T\Delta S^{\ddagger}$ with negative values for the natural substrate and non-substrate transition states (Table 1). The relative thermodynamics and kinetics of the reaction mechanism studied at B3LYP/6-31++G(d,p):AMBER level of theory for the enzyme—substrate complexes considered are presented in Table 1, all values are computed relative to the sum of separated reactants in kcal mol^{-1}. The large negative values of the $T\Delta S^{\ddagger}$ indicates that the mechanistic process is largely driven by entropy contributions for both the HIV-1 PR B and C-SA—peptide complexes and based on accepted notion a more negative entropy indicates greater movement restrictions for the inhibitor in the active site.[62, 89] This is due to steric restrictions and strong non-covalent enzyme—inhibitor interactions for certain parts of the ligand[62, 89] and could also apply to substrate interaction at the active site. Additionally, from the calculated results (Table 1), it is

observed 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, free energy profiles were computed for both scissile and non-scissile regions of the Gag and Gag-Pol polyprotein segments (Tables 1, S1 and Figure 1).

It is interesting to note that the scissile bond for MA-CA systems behaves well and gives the lowest activation energies (compared to the non-scissile bonds) as expected. This is not the case for the RH-IN systems, suggesting the model substrates are too short for proper recognition.

3.2 Substrate specificity and analysis of the complexes

Hydrogen bond (HB) interactions between ligand side chains and enzyme residues are a large factor determining the specificity of the complex "fit",[69] hence the HB interactions was considered for the MA-CA and RH-IN complexes in this study. It should be noted that both subtype B and C-SA complexes with MA-CA of the scissile bond exhibited approximately 22-21 HB interactions before optimization and 20-19 HB interactions after the calculation, respectively. RH-IN complexed with both subtypes B/C-SA PR has approximately 21-20 HB interactions before calculation and 17-16 HB interactions after calculation, respectively. As shown in Table 1 the MA-CA scissile bond cleaves faster compared to the RH-IN scissile bond, which can imply that the more the HB interactions the more the reactivity, and the lower the activation energy.

The subtype B—MA-CA natural substrate complexes have activation free energy differences ($\Delta\Delta G^{\ddagger}$) of approximately of 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.*[19] where HIV-1 substrates shows 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 non-scissile sequence with Val at the P2′ position, a higher activation free energy ($\Delta G^{\ddagger}$) 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 non-scissile bond with small residues (Ser*Gln) has a better affinity. Although, subtype B PR correctly recognizes and cleaves the non-scissile peptide bonds, the effect of the Pro side chain in Asn-Tyr-Pro*Ile-Val-Gln sequence gave a $\Delta\Delta G^{\ddagger}$ of 5.19 kcal mol^{-1} compared with the scissile bond domain (Gln-Val-Ser*Gln-Asn-Tyr). It is quite fascinating to note that the calculated $\Delta G^{\ddagger}$ (15.22 kcal mol^{-1}) for 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.[56]

For subtype C-SA—MA-CA complex, a $\Delta G^{\ddagger}$ of 13.49 kcal mol^{-1} was obtained for the cleavage of the scissile bond (Gln-Asn-**Tyr*Pro**-Ile-Val), with a decreased $\Delta\Delta G^{\ddagger}$ of -1.73 kcal mol^{-1} in comparison with the subtype B—MA-CA complex. Based on previous knowledge[47, 49] a higher activation energy has been associated with a diminished hydrolytic ability and perhaps an increase in potential bioactivity. The rate of hydrolysis for subtype C-SA—MA-CA is faster and this could be as a result of the mutations that occur in the protease, which also contributes to the less effective inhibition of the FDA approved drugs.[61, 90] This lower $\Delta G^{\ddagger}$ could also be related to the HIV-1 PR preference for hydrophobic amino acid side chains (Tyr*Pro) at the P1 and P1′ position of the natural target. Furthermore, the non-scissile bond domain that contain Pro*Ile has a $\Delta G^{\ddagger}$ 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).

Table 1: Relative thermodynamic and kinetic parameters for the one-step concerted acyclic TSs involved in the in the cleavage mechanism of the P3-P3' natural substrates for scissile and non-scissile bonds by HIV-1 PRs using ONIOM [B3LYP/6-31++G(d,p):AMBER]

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

Values are reported in kcal mol^{-1} relative to the sum of separated reactants.
The peptide sequences in bold are the natural substrates 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. All the 3D structures of the enzyme—substrate complexes considered are provided with the supporting information (Gaussian input file format).

Presented in Table 1 is the relative thermodynamic parameters for the enzyme—RH-IN catalysis, it was noted that both proteases recognize and cleave this natural substrate. For cleavage of subtype B—RH-IN natural substrate complex, a $\Delta G^{\ddagger}$ of 19.40 kcal mol^{-1} was obtained for the scissile bond, which is 0.07 kcal mol^{-1} higher than the non-scissile domain considered (Ile-Leu-Phe*Leu-Asp-Gly). This seems feasible as the presence of Asp amino acid unit at the P2′ position, which is more exposed to the active site and its ability to form a salt bridge, may contribute to its higher affinity to cleave the peptide. Similarly, subtype C-SA—RH-IN natural substrate system was observed to follow this trend with $\Delta\Delta G^{\ddagger}$ difference of -0.06 and 2.71 kcal mol^{-1} compared to the non-scissile peptide sequence Ile-Leu-Phe*Leu-Asp-Gly and Gly-Ile-Arg*Lys-Ile-Leu, respectively (Table 1). For both subtypes, HIV-1 PR complexed with the non-scissile domain Gly-Ile-Arg*Lys-Ile-Leu, the activation free energies of 22.00 and 20.74 kcal mol^{-1} were obtained, respectively.

The results for the polypeptide sequences reveal that subtype C-SA has an increase rate of hydrolysis compared to subtype B. It is worth mentioning that, depending on the amino acid sequence, experimental $\Delta G^{\ddagger}$ values between 14.86 and 21.03 kcal·mol^{-1} were reported in literature[13, 55, 57, 91-95] for the hydrolysis of peptides (including natural substrates) by HIV-1 PR and all the calculated activation free energies reported herein (Table 1) are in reasonable agreement with these experimental values. Specifically, the calculated $\Delta G^{\ddagger}$ of 18.03 and 17.97 kcal mol^{-1} for the cleavage of HIV-1 PR subtype C-SA—RH-IN scissile domain and Ile-Leu-Phe*Leu-Asp-Gly non-scissile region, respectively, is in agreement with observed experimental $\Delta G^{\ddagger}$ values of 18.24[13] and 17.60[57] kcal mol^{-1} for RH-IN hydrolysis by HIV-1 PR.

In a related theoretical investigation by Perez *et al.*,[96] it was concluded that many non-scissile peptide sequences, despite their high affinity do not cleave because they are suppressed inside the polyproteins.[96] Also, the recognition of substrate by HIV-1 PR is based on the geometric specificity of the exposed Gag and Gag-Pol polyprotein precursors.[19, 96] However, in this present work whereby the non-scissile domains were modelled in proximity with the PRs, they did not necessarily yield higher activation free energies compared to the natural substrates with the default scissile bonds (Table 1).

These results suggested that the model RH-IN system is too short for effective recognition. It was decided to study a slightly longer P5-P5' model substrate. A recent (experimental and

computational) study[97] also suggested that an increase in the peptide sequence has a substantial effect on the protease selectivity of cleavage sites.

The relative thermodynamic parameters for the HIV-1 PRs—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 non-scissile bonds. A $\Delta G^{\ddagger}$ of 19.69 kcal mol^{-1} was observed for the scissile bond cleavage of subtype B—RH-IN natural substrate complex, which is 0.11 kcal mol^{-1} lower than the non-scissile domain considered (Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp). Likewise, subtype C-SA—RH-IN natural substrate system was observed to follow this trend with $\Delta\Delta G^{\ddagger}$ difference of 0.10 and 2.70 kcal mol^{-1} compared to the non-scissile peptide sequence Arg-Lys-Ile-Leu-Phe*Leu-Asp-Gly-Ile-Asp and Ser-Ala-Gly-Ile-Arg*Lys-Ile-Leu-Phe-Leu, respectively (Table 2).

Table 3: Relative thermodynamic and kinetic parameters for the one-step concerted acyclic TSs involved in the in the cleavage mechanism of the P5-P5' natural substrates for scissile and non-scissile bonds by HIV-1 PRs using ONIOM [B3LYP/6-31++G(d,p):AMBER]

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

Values are reported in kcal mol^{-1} relative to the sum of separated reactants.
The peptide sequences in bold are the natural substrates 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. All the 3D structures of the enzyme—substrate complexes considered are provided with the supporting information (Gaussian input file format).

For both proteases, HIV-1 PR complexed with the non-scissile 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 for 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 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. 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 the amino acids residues in the substrate polypeptide sequence. The reaction mechanism follows a concerted acyclic TS model leading to product and activation free energies in reasonable agreement with experimental values for peptide (including natural substrate sequence) hydrolysis by HIV-1 PR. Substrates P5-P5' were required for suitable discrimination between the activation energies of scissile versus non-scissile bonds.

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 inhibits the subtype B better than the subtype C-SA PR as reported in our previous work.[61] Based on accepted theory that the HIV-1 PR inhibitors mimic the transition state of the enzyme-catalysed reaction,[20] 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 based on its mechanism of reaction.

Competing interests

The authors declare that they have no competing interests.

Supporting Information

The PDB formats of the ONIOM output files for the TS structures of the enzyme—substrate complexes and are available here.

Acknowledgements

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 Centre for High Performance Computing (www.chpc.ac.za) for computational resources.

References

[1] Feinberg, M. B. (1996) Changing the natural history of HIV disease, *The Lancet 348*, 239-246.

[2] Shehu-Xhilaga, M., and Oelrichs, R. (2009) Basic HIV virology, *HIV Management in Australasia*, 9-18.

[3] Hsu, I.-N., Delbaere, L. T., James, M. N., and Hofmann, T. (1977) Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 Å and sequence homology with porcine pepsin, *Nature 266*, 140.

[4] 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.

[5] 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.

[6] Weber, I. T., Cavanaugh, D. S., and Harrison, R. W. (1998) Models of HIV-1 protease with peptides representing its natural substrates, *Journal of Molecular Structure: THEOCHEM 423*, 1-12.

[7] 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.

[8] Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. (1985) Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene, *Science 229*, 1402-1405.

[9] Stanton, R. V., Hartsough, D. S., and Merz Jr, K. M. (1993) Calculation of solvation free energies using a density functional/molecular dynamics coupled potential, *The Journal of Physical Chemistry 97*, 11868-11870.

[10] 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.

[11] Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, *Nature 337*, 615.

[12] Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., and Geoghegan, K. F. (1989) X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes, *Nature 342*, 299.

[13] 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.

[14] Wlodawer, A., and Erickson, J. W. (1993) Structure-based inhibitors of HIV-1 protease, *Annual Review of Biochemistry 62*, 543-585.

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

[16] Lawal, M., Sanusi, Z., Govender, T., Maguire, G., Honarparvar, B., and Kruger, H. (2018) From recognition to reaction mechanism: an overview on the interactions between HIV-1 protease and its natural targets, *Current Medicinal Chemistry*.

[17] 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.
[18] Tözsér, J., Bláha, I., Copeland, T. D., Wondrak, E. M., and Oroszlan, S. (1991) Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins, *Febs Letters 281*, 77-80.
[19] Boross, P., Bagossi, P., Copeland, T. D., Oroszlan, S., Louis, J. M., and Tözsér, J. (1999) Effect of substrate residues on the P2′ preference of retroviral proteinases, *European Journal of Biochemistry 264*, 921-929.
[20] 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.
[21] Antonov, V. K., Ginodman, L. M., Rumsh, L. D., Kapitannikov, Y. V., Barshevskaya, T. N., Yavashev, L. P., Gurova, A. G., and Volkova, L. I. (1981) Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange, *The FEBS Journal 117*, 195-200.
[22] 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.
[23] 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.
[24] 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.
[25] Antonov, V., Ginodman, L., Kapitannikov, Y. V., Barshevskaya, T., Gurova, A., and Rumsh, L. (1978) Mechanism of pepsin catalysis: General base catalysis by the active-site carboxylate ion, *Febs Letters 88*, 87-90.
[26] 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.
[27] Polgar, L. (1989) *Mechanisms of protease action*, CRC press.
[28] Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Schneider, J., Kent, S., and Wlodawer, A. (1990) X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor, *Proceedings of the National Academy of Sciences 87*, 8805-8809.
[29] Rodriguez, E. J., Angeles, T. S., and 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*, 12380-12385.
[30] Bott, R., Subramanian, E., and Davies, D. (1982) Three-dimensional structure of the complex of the Rhizopus chinensis carboxyl proteinase and pepstatin at 2.5. ANG. resolution, *Biochemistry 21*, 6956-6962.
[31] Suguna, K. a., Padlan, E. A., Smith, C. W., Carlson, W. D., and Davies, D. R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action, *Proceedings of the National Academy of Sciences 84*, 7009-7013.

[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] Tripathi, A., Fornabaio, M., Spyrakis, F., Mozzarelli, A., Cozzini, P., and Kellogg, G. E. (2007) Complexity in Modeling and Understanding Protonation States: Computational Titration of HIV-1-Protease–Inhibitor Complexes, *Chemistry and Biodiversity 4*, 2564-2577.
[38] Brik, A., and Wong, C.-H. (2003) HIV-1 protease: mechanism and drug discovery, *Organic & Biomolecular Chemistry 1*, 5-14.
[39] Piana, S., and 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*, 26-36.
[40] Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values, *Proteins: Structure, Function, and Bioinformatics 61*, 704-721.
[41] Honarparvar, B., Govender, T., Maguire, G. E., Soliman, M. E., and Kruger, H. G. (2013) Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity, *Chemical Review 114*, 493-537.
[42] Goldblum, A., Glick, M., and Rayan, A. (1993) Determining proton positions in an enzyme-inhibitor complex is a first step for theoretical mechanistic studies of aspartic proteinases, *Theoretical Chemistry Account 85*, 231-247.
[43] Tropsha, A., and Hermans, J. (1992) Application of free energy simulations to the binding of a transition-state-analogue inhibitor to HTV protease, *Protein Engineering, Design and Selection 5*, 29-33.
[44] Harrison, R. W., and Weber, I. T. (1994) Molecular dynamics simulations of HIV-1 protease with peptide substrate, *Protein Engineering, Design and Selection 7*, 1353-1363.
[45] Beveridge, A., and Heywood, G. (1993) A quantum mechanical study of the active site of aspartic proteinases, *Biochemistry 32*, 3325-3333.
[46] 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.
[47] Lawal, M. M., Govender, T., Maguire, G. E., 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.

[48] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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.

[49] Lawal, M. M., Govender, T., Maguire, G. E., 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*.

[50] Becke, A. D. (1993) Density-functional thermochemistry. III. The role of exact exchange, *The Journal of Chemical Physics 98*, 5648-5652.

[51] Rassolov, V. A., Pople, J. A., Ratner, M. A., and Windus, T. L. (1998) 6-31G* basis set for atoms K through Zn, *The Journal of Chemical Physics 109*, 1223-1229.

[52] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G* basis set for third-row atoms, *Journal Computational Chemistry 22*, 976-984.

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

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

[55] Kipp, D. R., Silva, R. G., and Schramm, V. L. (2011) Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease, *Journal of the American Chemical Society 133*, 19358-19361.

[56] Maschera, B., Darby, G., Palú, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and 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*, 33231-33235.

[57] Rodriguez, E. J., Debouck, C., Deckman, I. C., Abu-Soud, H., Raushel, F. M., and Meek, T. D. (1993) Inhibitor binding to the Phe53Trp mutant of HIV-1 protease promotes conformational changes detectable by spectrofluorometry, *Biochemistry 32*, 3557-3563.

[58] Lawal, M. M., Sanusi, Z. K., Govender, T., Tolufashe, G. F., Maguire, G. E., Honarparvar, B., and 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*, 1-9.

[59] Kempf, D. J., Norbeck, D. W., Codacovi, L., Wang, X. C., Kohlbrenner, W. E., Wideburg, N. E., Paul, D. A., Knigge, M. F., and Vasavanonda, S. (1990) Structure-based, C2 symmetric inhibitors of HIV protease, *Journal of Medicinal Chemistry 33*, 2687-2689.

[60] Naicker, P., Achilonu, I., Fanucchi, S., Fernandes, M., Ibrahim, M. A., Dirr, H. W., Soliman, M. E., and 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*, 1370-1380.

[61] Sanusi, Z., Govender, T., Maguire, G., 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.

[62] Sanusi, Z. K., Govender, T., Maguire, G. E., 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.

[63] Field, M. J., Bash, P. A., and Karplus, M. (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, *Journal of Computational Chemistry 11*, 700-733.

[64] Heinrich, N. (1992) KB Lipkowitz, DB Boyd (Eds.): Reviews in Computational Chemistry, VCH Verlagsgesellschaft, Weinheim, New York, 1990, ISBN 3-527-27845-1, 419 Seiten, Preis: DM 176,–, *Reports of the Bunsen Society for Physical Chemistry 96*, 224-224.

[65] Wei, D., and Salahub, D. (1994) A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution, *Chemical physics letters 224*, 291-296.

[66] 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.

[67] Vreven, T., and Morokuma, K. (2000) On the application of the IMOMO (integrated molecular orbital+ molecular orbital) method, *Journal of Computational Chemistry 21*, 1419-1432.

[68] 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.

[69] 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.

[70] 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.

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

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

[73] 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.

[74] Costa, M. G. S., Benetti-Barbosa, T. G., Desdouits, N., Blondel, A., Bisch, P. M., Pascutti, P. G., and Batista, P. R. (2014) Impact of M36I polymorphism on the interaction of HIV-1 protease with its substrates: insights from molecular dynamics, *BMC Genomics 15 Suppl 7*, S5.

[75] Sadiq, S. K., Wright, D., Watson, S. J., Zasada, S. J., Stoica, I., and Coveney, P. V. (2008) Automated Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases, *Journal of Chemical Information and Modeling 48*, 1909-1919.

[76] Ozen, A., Haliloglu, 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.
[77] DeLano, W. L. (2002) The PyMOL molecular graphics system.
[78] Dennington, R., Keith, T., and Millam, J. (2009) Semichem Inc, *Shawnee Mission KS, GaussView, Version 5*.
[79] Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins, *Journal of Molecular Biology 238*, 415-436.
[80] Ribeiro, A. J. M., Santos-Martins, D., Russo, N., Rarnos, M. J., and Fernandes, P. A. (2015) Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease, *ACS Catalysis 5*, 5617-5626.
[81] Lee C., Yang W., and R.G, P. (1988) Development of Colle-Salvetti correlation-energy formula into a functional of electron density, *Phsical Review B 37*, 785-789.
[82] 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.
[83] Vreven, T., and Morokuma, K. (2006) Hybrid methods: Oniom (qm: mm) and qm/mm, *Annual Reports in Computational Chemistry 2*, 35-51.
[84] Humbel, S., Sieber, S., and Morokuma, K. (1996) 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−, *The Journal of Chemical Physics 105*, 1959-1967.
[85] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *Journal of Computational Chemistry 25*, 1157-1174.
[86] Ochterski, J. W. (1999) Vibrational analysis in Gaussian, *help@ gaussian. com*.
[87] Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B., and Petersson, G. (2009) 09, Revision D. 01, Gaussian, *Inc., Wallingford, CT*.
[88] Ochterski, J. W. (2000) Thermochemistry in gaussian, *Gaussian Inc*, 1-19.
[89] Ryde, U. (2014) A fundamental view of enthalpy–entropy compensation, *Medchemcomm 5*, 1324-1336.
[90] Maseko, S. B., Padayachee, E., Govender, T., Sayed, Y., Kruger, G., Maguire, G. E., and Lin, J. (2017) I36T↑ T mutation in South African subtype C (C-SA) HIV-1 protease significantly alters protease-drug interactions, *Biological Chemistry 398,* 1109-1117.
[91] Feher, A., Weber, I. T., Bagossi, P., Boross, P., Mahalingam, B., Louis, J. M., Copeland, T. D., Torshin, I. Y., Harrison, R. W., and Tozser, J. (2002) Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites, *European Journal of Biochemistry 269*, 4114-4120.
[92] Tozser, J., Bagossi, P., Weber, I. T., Copeland, T. D., and Oroszlan, S. (1996) Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases, *Journal of Biological Chemistry 271*, 6781-6788.
[93] Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D., and 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*, 33231-33235.

[94] Hyland, L. J., Tomaszek Jr, T. A., Roberts, G. D., Carr, S. A., Magaard, V. W., Bryan, H. L., Fakhoury, S. A., Moore, M. L., Minnich, M. D., and 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.
[95] Boross, P., Bagossi, P., Copeland, T. D., Oroszlan, S., Louis, J. M., and Tozser, J. (1999) Effect of substrate residues on the P2 ' preference of retroviral proteinases, *European Journal of Biochemistry 264*, 921-929.
[96] 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.
[97] Laco, G. S. (2017) Retroviral proteases: correlating substrate recognition with both selected and native inhibitor resistance, *Journal of Molecular Biochemistry 6*.

Supporting Information for Chapter Five

A Computational 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.

The nine recognition sequences cleaved by HIV-1 PR

Table S4: Recognition sequences for natural substrates cleaved by the HIV-1 protease.[60, 61]

Peptide sequences cleavage domain	Cleavage domain
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	
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.[60, 61]

The 3D structures of the enzyme—substrate complexes.

Sub_recog_pdb_structures.zip

CHAPTER SIX

CONCLUSION

Research on the Human Immunodeficiency Virus type 1 (HIV-1) is one of the most explored viral infection across the globe. The investigations of which has expanded to the molecular understanding of the virus aimed at drug design[1] to manage the spread of the virus within its host. Since the discovery of azidothymidine (AZT) often referred to as zidovudine, in the 1980s,[2] the use of drugs inhibiting certain developmental stage in the life cycle of the HI virus has remained a renowned method of therapeutic management.

The design of drugs inhibiting specific enzymes necessary for the replication of the virus formed part of the mainstream research for over two decades. Amongst these targeted 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.[3-6] The mechanistic action of this PR on the cleavage of these polyproteins has been a subject of research for the past three decades. Despite tremendous contributions from across the globe on understanding the recognition and reaction mechanism of HIV-1 PR with respect to its natural substrates, there are important gaps yet to be addressed by researchers.

To put this section into perspective, a brief overview of the dissertation is presented. Provided in the introductory chapter is a general discourse of the topic that provided an overview of the necessary background for the rest of the thesis (Chapter one). A detailed literature review (from the year 2000) on the recognition pattern and hydrolysis of natural substrate by HIV-1 PR was presented in chapter two. Using quantum mechanics methods, a representative HIV-1 PR—substrate model was used to investigate the reaction mechanism (Chapter three) and a very interesting outcome was obtained. This study was then widened to understanding the cleavage mechanism of HIV-1 PR on its substrates at the molecular level using an enzyme—substrate system (Chapter four) and a hybrid QM/MM method. The recognition mechanism of asymmetric natural substrate by homodimeric HIV-1 PR was also investigated (Chapter five) using a hybrid QM/MM method. The entire work was thus summarized in this section (Chapter six) to provide overall conclusion on the present study.

In this present work, which was initiated by an introductory background study (Chapter one) and a comprehensive literature review (Chapter two) of both theoretical and experimental investigations, it was concluded that HIV-1 PR recognizes its natural substrate through; substrate

modulation, conserved substrate shape, interdependence conformational adaptability of both the enzyme and their natural targets.[7-10] A notable contrasting opinion from computational simulations[11] 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.[11] The reaction mechanism of HIV-1 PR has been more often investigated as a stepwise general acid-base rather than a concerted process. However, the stepwise reaction pathway of HIV-1 PR—substrate seems to be characterized by controversies including the exact rate-limiting step.[12-14] Also, the protonation pattern of the catalytic aspartates present in the HIV-1 PR active site, responsible for the hydrolytic cleavage of substrates remains a subject of debate and was addressed herein.

The identifiable gaps from the literature review prompted the design of a one-step concerted general acid-base mechanism of HIV-1 PR—substrate system in which the process occurs *via* cyclic transition state (TS) as proposed earlier from experiment.[15] In the light of literature support for similar six-membered ring transition structures observed theoretically[16-23] a one-step concerted chemical process was explored using computational approach with the aim of providing an alternative insight on the catalytic mechanism of HIV-1 PR and inhibitor design. In other words, the recognition pattern and the reaction mechanism of natural substrates by HIV-1 PRs was investigated by means of concerted TS modelling after thorough re-examination of the subject from relevant literature.

The first computational study was dedicated to understanding the enzymatic mechanism of HIV-1 PR on natural substrate and derivatives using both DFT (M06-2X) and HF methods for a small model system (Chapter three). A previously studied one-step six-membered cyclic TS structure[16-23] was mimicked. The present study addressed the HIV-1 PR—natural substrate mechanistic pathway through the estimation of thermodynamics, kinetics, and other quantum chemical characteristics. The investigation enabled us to deduce a correlation between the calculated activation free energies and C—N force constants of the N-substituted amide derivatives thereby, buttressing the relevance of bond strength in overcoming activation barrier. In other words, the higher the activation free energy, the higher the force constant value which also correlates to higher C—N bond strength in each substrate. The models allowed the examination of N-substituted scissile bond derivatives which provided us with a new potential lead (methylthio amide) for HIV-1 PR inhibition.

In the second investigation, a new (theoretical) perspective on three possible concerted general acid-base mechanisms was observed using a QM/MM (ONIOM) model (Chapter four). For the first time, a concerted one-step linear mechanism for the hydrolysis of the natural substrate in the HIV PR enzyme was observed. The mechanism that provides marginally the lowest activation barrier involves an acyclic TS model with one water molecule at the HIV-1 PR active site. A two-water model involving cyclic TS structure was also proposed for a favourable catalysis of substrates and should be pursued experimentally in subsequent research. The 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. These models also provide new 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 models could be applied to related homodimeric protease and perhaps other enzymatic processes.

Natural substrates' recognition pattern by HIV-1 PRs was investigated last through the concerted acyclic TS modelling using ONIOM method (Chapter five). It was observed that the recognition specificity of the enzymes depends on the sequence of the amino-acid residues within the natural substrate. The studied HIV-1 PR subtypes (B and C-SA) recognize and cleave at both scissile and non-scissile regions of the natural substrate sequences and maintaining preferential specificity for the scissile bonds with characteristic lower activation free energies.

Studies on HIV-1 PR are quite diverse and theoretical approaches had played a vital role in determining the recognition pattern and enzymatic mechanism of HIV-1 PR and its natural targets. 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. In this present study, we observed a substantial difference in the energy barriers of both the small model system and the entire enzyme—substrate complex.

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 computational efforts should explore the applications of sophisticated computational techniques aimed at revealing the overall free energy landscape of the HIV-1 PR potential energy surface. 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.

References

[1] 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*, pp 203-241, Springer.

[2] Yarchoan, R., Mitsuya, H., Myers, C. E., and Broder, S. (1989) Clinical pharmacology of 3'-azido-2', 3'-dideoxythymidine (zidovudine) and related dideoxynucleosides, *New England Journal of Medicine 321*, 726-738.

[3] 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.

[4] 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.

[5] 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*, 280-283.

[6] 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.

[7] 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 protease, *Journal of Molecular Biology 301*, 1207-1220.

[8] 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.

[9] 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.

[10] Ozer, N., Haliloglu, T., and Schiffer, C. A. (2006) Substrate specificity in HIV-1 protease by a biased sequence search method, *Proteins: Structure, Function, and Bioinformatics 64*, 444-456.

[11] 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.

[12] Calixto, A. R., Ramos, M. J., and Fernandes, P. A. (2017) Influence of Frozen Residues on the Exploration of the PES of Enzyme Reaction Mechanisms, *Journal of Chemical Theory and Computation 13*, 5486-5495.

[13] Ribeiro, A. J. M., Santos-Martins, D., Russo, N., Rarnos, M. J., and Fernandes, P. A. (2015) Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease, *ACS Catalysis 5*, 5617-5626.

[14] 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.
[15] 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.
[16] Kruger, H. G. (2002) Ab initio mechanistic study of the protection of alcohols and amines with anhydrides, *Journal of Molecular Structure: Theochem 577*, 281-285.
[17] 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[5.4.0.0(2,6).0(3,10).0(5,9)] undecane-8,11-lactam, *Journal of Molecular Structure- Theochem 672*, 119-125.
[18] 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.
[19] Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E. S., Honarparvar, B., Parboosing, R., Naidoo, A., Arvidsson, P. I., Sayed, Y., Govender, P., Maguire, G. E. M., Kruger, H. G., and Govender, T. (2011) Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: A hybrid 2D NMR and docking/QM/MM/MD approach, *European Journal Medicinal Chemistry 46*, 3976-3985.
[20] 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.
[21] 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.
[22] Lawal, M. M., Govender, T., Maguire, G. E., 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.
[23] Fakhar, Z., Govender, T., Lamichhane, G., Maguire, G. E., 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.