

โครงสร้างและพลวัตของเฮชไอวี-1 อินทิเกรต:  
แบบจำลองเชิงโมเลกุลและการจำลองพลศาสตร์เชิงโมเลกุล

นางสาวอัญชรา วิจิตรโกสุม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาเคมี ภาควิชาเคมี

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2547

ISBN 974-53-1050-6

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I22035473

STRUCTURE AND DYNAMIC OF HIV-1 INTEGRASE:  
MOLECULAR MODELING AND MOLECULAR DYNAMICS SIMULATIONS

Miss Atchara Wijitkosoom

A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy in Chemistry

Department of Chemistry

Faculty of Science

Chulalongkorn University

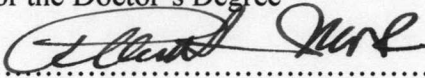
Academic year 2004

ISBN 974-53-1050-6

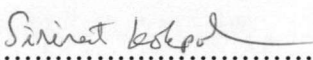
Thesis Title                      Structure and Dynamic of HIV-1 Integrase: Molecular Modeling  
and Molecular Dynamics Simulations  
By                                      Miss Atchara Wijitkosoom  
Filed of study                      Chemistry  
Thesis Advisor                      Associate Professor Supot Hannongbua, Ph. D.  
Thesis Co-advisor                      Associate Professor Vudhichai Parasuk, Ph. D.  
Thesis Co-advisor                      Professor Thanh N. Truong, Ph. D.

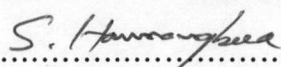
---

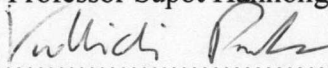
Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment  
of the Requirements for the Doctor's Degree

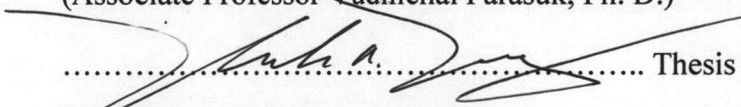
.....  ..... Dean of the Faculty of Science  
(Professor Piamsak Menasveta, Ph. D.)

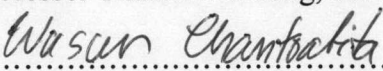
THESIS COMMITTEE

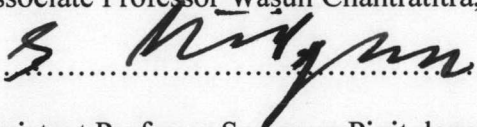
.....  ..... Chairman  
(Associate Professor Sirirat Kokpol, Ph. D.)

.....  ..... Thesis Advisor  
(Associate Professor Supot Hannongbua, Ph. D.)

.....  ..... Thesis Co-advisor  
(Associate Professor Vudhichai Parasuk, Ph. D.)

.....  ..... Thesis Co-advisor  
(Professor Thanh N. Truong, Ph. D.)

.....  ..... Member  
(Associate Professor Wasun Chantratitra, Ph. D.)

.....  ..... Member  
(Assistant Professor Surapong Pinitglang, Ph. D.)

อัจฉรา วิจิตรโกสม : โครงสร้างและพลวัตของเอชไอวี-1 อินทิเกรต: แบบจำลองเชิงโมเลกุล และการจำลองพลศาสตร์เชิงโมเลกุล. (STRUCTURE AND DYNAMIC OF HIV-1 INTEGRASE: MOLECULAR MODELING AND MOLECULAR DYNAMICS SIMULATIONS) อ. ที่ปรึกษา : รศ. ดร. สุพจน์ หารหนองบัว, อ.ที่ปรึกษาร่วม : รศ. ดร. วุฒิชัย พาราสุข, Professor Thanh N. Truong, 158 หน้า. ISBN 974-53-1050-6.

ได้สร้างแบบจำลองโครงสร้างที่สมบูรณ์ของเอชไอวี-1 อินทิเกรตซึ่งเป็นหนึ่งในเอนไซม์เป้าหมายสำหรับการออกแบบยาต้านเอชไอวี โดยใช้ข้อมูลทางการทดลองของส่วนประกอบของเอนไซม์ ทั้งนี้ได้เลือกใช้โครงสร้างของเอนไซม์แบบสองโดเมน เพื่อให้ได้โครงสร้างสมบูรณ์ที่น่าเชื่อถือ โดยเฉพาะอย่างยิ่งที่บริเวณเชื่อมต่อระหว่างคอร์กับเอ็น-เทอร์มินัลโดเมน (คอร์-เอ็น) และคอร์กับซี-เทอร์มินัลโดเมน (คอร์-ซี) ในการศึกษานี้ได้คำนวณโดยใช้วิธีการจำลองพลศาสตร์เชิงโมเลกุลเป็นเวลาสองนาโนวินาที สำหรับโครงสร้างสมบูรณ์ของเอชไอวี-1 อินทิเกรตทั้งกรณีที่มีและไม่มีไอออนแมกนีเซียมในบริเวณเร่งของคาตาไลติกคอร์โดเมน นอกจากนี้ยังได้คำนวณสมบัติทางโครงสร้างและพลศาสตร์ของโครงสร้างหนึ่งโดเมน (คอร์) และ โครงสร้างสองโดเมน (คอร์-เอ็น และคอร์-ซี) เพื่อศึกษาอิทธิพลของปลายเทอร์มินัลที่มีต่อคาตาไลติกคอร์โดเมน ผลการศึกษาแสดงให้เห็นว่าบริเวณที่หายไปจากการทดลองในโครงสร้างทุกแบบมีความยืดหยุ่นสูง และเมื่อเปรียบเทียบกับในระหว่างสามโดเมนพบว่า ซี-เทอร์มินัลโดเมนมีความยืดหยุ่นมากที่สุด ซึ่งความยืดหยุ่นที่สูงมากนี้ มีความเป็นไปได้ที่จะช่วยในกระบวนการเข้าจับของดีเอ็นเอในปฏิกิริยาอินทิเกรชัน นอกจากนี้ยังพบว่าไอออนโลหะมีอิทธิพลต่อการจัดตัวและคอนฟอร์เมชันของคีย์เรสิดิวในบริเวณเร่ง ซึ่งแสดงในรูปของการเปลี่ยนแปลงระยะทางและมุมระหว่างเรสิดิวบริเวณเร่งนั้นกับ ไอออนโลหะ นอกจากนี้ ยังได้สร้างแบบจำลองโครงสร้างสมบูรณ์แบบไดเมอร์และเตตราเมอร์ของเอชไอวี-1 อินทิเกรตที่จับกับดีเอ็นเอบนพื้นฐานของข้อมูลทางการทดลองที่มีอยู่

ภาควิชา.....เคมี.....ลายมือชื่อนิสิต.....  
 สาขาวิชา.....เคมี.....ลายมือชื่ออาจารย์ที่ปรึกษา.....  
 ปีการศึกษา.....2547.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....  
 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

## 4473850623 : MAJOR CHEMISTRY

KEY WORD: HIV-1 INTEGRASE / FULL-LENGTH / MOLECULAR DYNAMICS / STRUCTURE

ATCHARA WIJITKOSOOM : STRUCTURE AND DYNAMIC OF HIV-1 INTEGRASE:  
MOLECULAR MODELING AND MOLECULAR DYNAMICS SIMULATIONS. THESIS  
ADVISOR : ASSOC. PROF. DR. SUPOT HANNONGBUA, THESIS COADVISOR :  
ASSOC. PROF. DR. VUDHICHAIR PARASUK, PROFESSOR DR. THANH N. TRUONG,  
158 pp. ISBN 974-53-1050-6.

The complete full-length structure of HIV-1 integrase, a promising target for the design of anti-AIDS drugs, was theoretically modeled from experimental data of its fragments. The two experimental structures of two-domain fragment were used to carefully build the reliable full-length structure, particularly the linkages between core and N-terminal domains (CORE-N), and core and C-terminal domains (CORE-C). 2-ns molecular dynamics simulations were performed for the full-length HIV-1 integrase model in two cases, namely with and without a  $Mg^{2+}$  ion in the active site of the catalytic core domain. The structural and dynamical properties of the one-domain fragment, CORE, and the two-domain fragments, CORE-N and CORE-C, were also calculated in order to investigate effect of the terminal end on the catalytic core domain. The results show high flexibility in the experimentally missing region of the HIV-1 structure in all cases. Among the three domains, the C-terminal has the largest flexibility. Such flexibility was supposed to facilitate in the DNA binding process during the integration reaction. The metal ion has significant effects on the orientation and conformation of key residues in the active site as indicated by substantial changes in the distances and angles between the residues and ion in the active site region. A dimer and tetramer full-length HIV-1 IN complexed with DNA was successfully modeled herein based on the available experimental data.

Department..... Chemistry..... Student's signature..... *A. Wijitkoom*  
Field of study..... Chemistry..... Advisor's signature..... *S. Hannongbua*  
Academic year..... 2004 ..... Co-advisor's signature..... *Vudhichai Parasuk*  
Co-advisor's signature..... *Thanh N. Truong*

## ACKNOWLEDGEMENTS

First, all affections gratitude is acknowledged to Wijitkosoom family, my parents, Mr. Wirote and Mrs. Thatsanee for their understandings, supports and encouragements throughout my entire life, and my siblings, Mr. Wisit, Miss Saowanee and Miss Jadesadaporn for joining me during the entire student life.

This thesis was completely finished with the excellent helps from my thesis advisor, Assoc. Prof. Dr. Supot Hannongbua, who always gives me his guiding, advising and encouraging for all my graduate study in Japan, USA and Thailand. This is also to my co-advisor Assoc. Prof. Dr. Vudhichai Parasuk. I would like to thank Prof. Dr. Thanh N. Truong at the University of Utah, my thesis co-advisor for his suggestion and support during my stay in Utah. This is also to Assist. Prof. Dr. Thomas Cheatham for his kind AMBER assistance during my study in Utah. Thank Luca, Podtelezhnikov and their groups for providing the HIV-1 IN full-length models.

I would also like to express my special thanks to Assist. Prof. Dr. Somsak Tonmunphean for his understandings and supports in any situation. I gratefully acknowledge Assist. Prof. Dr. Pornthep Sompornpisut, my M. Sc. degree thesis advisor, who gives me the basic computer skill starting from zero. I am truly recognized Prof. Akira Miyamoto at Tohoku University, Japan for encouraging me to be stronger in doing research during my stay in Sendai, Japan. I am very appreciating to Assoc. Prof. Dr. Sirirat Kokpol, Assoc. Prof. Dr. Wasun Chantratitra and Assist. Prof. Dr. Surapong Pinitglang for their substantial advice as thesis committee.

I am gratefully thanks to Computational Chemistry Unit Cell (CCUC) at Department of Chemistry, Chulalongkorn University and the Center for High Performance Computing (CHPC) at the University of Utah for providing the computation time and all facilities.

Finally, this thesis was completely finished with the funding from the Royal Golden Jubilee (RGJ) Ph.D. program and the Thailand Research Fund (TRF) and Department of Chemistry, the University of Utah, Salt Lake City, Utah, USA.

# CONTENTS

	Pages
<b>ABSTRACT IN THAI</b> .....	iv
<b>ABSTRACT IN ENGLISH</b> .....	v
<b>ACKNOWLEDGEMENTS</b> .....	vi
<b>CONTENTS</b> .....	vii
<b>LIST OF TABLES</b> .....	xii
<b>LIST OF FIGURES</b> .....	xiv
<b>NOTATIONS</b> .....	xxvi
<b>CHAPTER 1 INTRODUCTION</b> .....	1
1.1 Research rationale.....	1
1.1.1 Molecular modeling: The most potentially method for drug design.....	1
1.1.2 HIV-1 integrase: The third target for AIDS therapy.....	2
1.1.3 Aims of this study.....	2
1.2 AIDS.....	3
1.2.1 Historical Outline.....	3
1.2.2 World epidemic situation.....	3
1.2.3 HIV/AIDS symptom.....	3
1.3 HIV.....	6
1.3.1 Historical Outline.....	6
1.3.2 Biology of HIV.....	7
1.3.3 Replication cycle.....	8
1.4 Treatment and FDA approved drugs.....	9
1.4.1 Historical Outline.....	9
1.4.2 Progress of inhibitors design for anti-retroviral drugs .....	9
1.5 HIV-1 integrase.....	13
1.5.1 Historical Outline.....	13
1.5.2 Structure and Function.....	13
1.5.2.1 N-terminal domain.....	13

	Pages
1.5.2.2 Catalytic core domain.....	14
1.5.2.3 C-terminal domain.....	15
1.5.3 Integration process.....	16
1.5.4 HIV-1 IN inhibitors.....	16
1.6 Interaction with DNA.....	21
1.6.1 Structure of DNA.....	21
1.6.2 Nucleotides: A building block of DNA.....	22
1.6.3 Double helical structure of DNA.....	23
1.6.4 HIV-1 IN-DNA interaction.....	26
1.6.4.1 Recognition of the viral and host DNA.....	26
1.6.4.2 Retroviral integration mechanism.....	27
1.6.4.3 Interaction with HIV-1 IN.....	27
1.7 Previous studies of the HIV-1 IN.....	28
1.7.1 Molecular dynamics simulation studies.....	28
1.7.2 Theoretical studies of the HIV-1 IN-DNA complex	34
1.8 Objectives.....	36
<b>CHAPTER 2 THEORY.....</b>	<b>37</b>
2.1 Molecular modeling.....	37
2.1.1 Homology modeling.....	37
2.1.1.1 Sequence alignment.....	38
2.1.1.2 Assignment of coordinates.....	40
2.1.2 Energy minimization.....	42
2.1.2.1 Steepest descents method.....	42
2.1.2.2 Conjugate gradient method.....	43
2.2 Molecular dynamics (MD) simulation.....	44
2.2.1 Theory.....	44
2.2.2 Potential function (force field).....	47
2.2.3 Periodic boundary condition.....	49
2.2.4 Potential cut-off for non-bonded interaction.....	49



	Pages
2.2.5 Treatment for long range interaction.....	49
2.3 Molecular docking.....	50
2.3.1 Gaussian density representation of protein shape...	52
2.3.2 Fourier expansion and coordinate operations.....	52
<b>CHAPTER 3 COMPUTATIONAL DETAILS.....</b>	<b>54</b>
3.1 Molecular modeling.....	54
3.1.1 Experiment structures: A theoretical point of view	54
3.1.1.1 The one-domain structures.....	54
3.1.1.1.1 N-terminal domain.....	54
3.1.1.1.2 Catalytic core domain.....	56
3.1.1.1.3 C-terminal domain.....	57
3.1.1.2 The two-domain fragment structures.....	58
3.1.1.2.1 Core connected to N-terminal domain.....	58
3.1.1.2.2 Core connected to C-terminal domain.....	59
3.1.1.3 Comparison of HIV-1 IN domain structures.....	65
3.1.1.3.1 N-terminal domain.....	65
3.1.1.3.2 Catalytic core domain.....	67
3.1.1.3.3 C-terminal domain.....	67
3.1.2 Homology modeling step.....	70
3.1.3 Modeling of the full-length structure.....	71
3.1.3.1 Full-length monomer.....	71
3.1.3.2 Full-length dimer and tetramer.....	73
3.2 MD simulations.....	74
3.2.1 Basic steps in the AMBER 7.....	74
3.2.2 Set up the MD simulation for the five systems.....	75
3.3 Molecular docking.....	79

	Pages
<b>CHAPTER 4 RESULTS AND DISCUSSION.....</b>	<b>81</b>
4.1 Molecular modeling.....	81
4.1.1 Homology modeling of the missing region.....	81
4.1.1.1 The one-domain structures.....	81
4.1.1.2 The two-domain structures.....	81
4.1.1.2.1 CORE-N.....	81
4.1.1.2.2 CORE-C.....	81
4.1.2 Modeling of the full-length HIV-1 IN.....	84
4.1.2.1 The full-length monomer.....	84
4.1.2.2 The full-length dimer.....	86
4.1.2.3 The full-length tetramer.....	87
4.1.3 Comparison of the modeling structure.....	88
4.1.3.1 The core domain.....	88
4.1.3.2 The N-terminal domain.....	90
4.1.3.3 The C-terminal domain.....	90
4.1.3.4 The two-domain structure.....	91
4.1.3.5 The full-length.....	92
4.2 MD simulations.....	96
4.2.1 The one-domain structure.....	96
4.2.1.1 CORE.....	96
4.2.2 The two-domain system.....	101
4.2.2.1 CORE-N.....	101
4.2.2.2 CORE-C.....	105
4.2.3 The three-domain system: The full-length system...	110
4.2.3.1 FULL.....	110
4.2.3.2 FULL+ION.....	113
4.2.4 Detailed analysis.....	117
4.2.4.1 Effect of divalent metal ion on the catalytic triad.....	121
4.2.4.2 Effect of the two end domains.....	122

	Pages
4.2.4.2.1 Effect on the structure.....	122
4.2.4.2.2 Effect on the mobility.....	123
4.2.4.3 Similar and dissimilar among the full- length model.....	126
4.3 Molecular docking.....	133
4.3.1 Dimer model bound to DNA.....	133
4.3.2 Tetramer model bound to DNA.....	133
4.3.3 HIV-1 IN-DNA interaction.....	134
4.4 Summary.....	138
4.4.1 Molecular modeling.....	138
4.4.2 MD simulations.....	138
4.4.3 Molecular docking.....	139
<b>CHAPTER 5 CONCLUSIONS.....</b>	<b>141</b>
5.1 Full-length model.....	141
5.2 Structural and dynamical properties.....	142
5.3 Effect of divalent metal ion.....	143
5.4 Effect of the two-end terminal domains.....	143
5.5 Dimer and tetramer full-length HIV-1 IN complexed with DNA.....	144
5.6 Suggestion for further works.....	145
<b>REFERENCES.....</b>	<b>146</b>
<b>CURRICULUM VITAE.....</b>	<b>158</b>

## LIST OF TABLES

		Pages
Table 1.1	Estimated number of people living with HIV/AIDS and number of deaths due to AIDS globally 1999 – 2003.	5
Table 1.2	Global summary of the HIV/AIDS (December 2003), the number is an average while the number in parenthesis is for range. The data is from UNAIDS 2004 report on the global AIDS epidemic.	5
Table 1.3	FDA approved anti-HIV drug.	11 - 12
Table 1.4	Structure, group and sites of action of the HIV-1 IN inhibitors.	18 - 19
Table 1.5	Selective historical review of the previous MD studies on the HIV-1 IN.	30 - 33
Table 3.1	Number of water molecules, cations, counter ions, water molecules and total atoms together with the dimension of simulation boxes of the seven simulations, CORE, CORE-N, CORE-C, FULL and FULL+ION.	61 - 64
Table 3.2	Missing residues in the two-domain structures, CORE-N and CORE-C, and the full-length (FULL).	73
Table 3.3	Number of water molecules, cations, counter ions, water molecules and total atoms together with the dimension of simulation boxes of the seven simulations, CORE, CORE-N, CORE-C, FULL+ION and FULL.	78
Table 4.1	Mean RMSD values for each comparison made between the domain in CORE, CORE-N, CORE-C and FULL with the individual domain in the FULL+ION structure.	119
Table 4.2	Distances ( $d1 - d4$ ) between the $Mg^{2+}$ and the carboxyl oxygen of Asp64 and Asp116.	120

Table 4.3	List of the amino acid residues which interact with the viral and host DNA from various studies.	136 - 137
-----------	--	-----------

## LIST OF FIGURES

		<b>Pages</b>
Figure 1.1	Stages of HIV/AIDS infection, acute, chronic (asymptomatic) and AIDS.	4
Figure 1.2	Structure of an HIV-1 virion and its component	6
Figure 1.3	Organization of HIV-1 genome and virion	7
Figure 1.4	The HIV infection cycle	8
Figure 1.5	Study of retroviral inhibitors design using various experimental and theoretical techniques.	10
Figure 1.6	Schematic representation of the full-length HIV-1 IN	13
Figure 1.7	N-terminal domain of the HIV-1 IN	14
Figure 1.8	Core domain of the HIV-1 IN	15
Figure 1.9	(a) Two conserved aspartate residue (Asp64, Asp116) and a glutamate residue (Glu152) coordinated with Mg <sup>2+</sup> , and (b) the proposed two ion mechanism	15
Figure 1.10	C-terminal domain of the HIV-1 IN	16
Figure 1.11	Integration process catalyzed by HIV-1 IN.	17
Figure 1.12	Chemical structures of some HIV-1 IN diketo acid inhibitors.	20
Figure 1.13	Structure of DNA.	22
Figure 1.14	The component of DNA (left) and structure of five bases, A, T, C, G and U (right), the hydrogen bonds formed by base pairs were also shown, the image was downloaded from the site: <a href="http://www.accessexcellence.org/RC/VL/GG/nucleotide2.html">http://www.accessexcellence.org/RC/VL/GG/nucleotide2.html</a> .	23

## LIST OF FIGURES (cont.)

		<b>Pages</b>
Figure 1.15	The double helical model of DNA by Watson and Crick. The picture was downloaded from Science photo library (by A. Barrington Brown).	24
Figure 1.16	Structure of the double helix DNA and its key features.	25
Figure 1.17	The structures of DNA in A, B and Z form.	25
Figure 1.18	Integration mechanism of the HIV-1 IN.	26
Figure 1.19	Core and C-terminal domain of HIV-1 IN complex with target and viral DNA from photo cross-linking study.	28
Figure 1.20	Proposed theoretical models for HIV-1 IN-DNA interaction.	35
Figure 2.1	Amino acid identity matrix.	38
Figure 2.2	Addition of an unknown sequence to the alignment, the unknown sequence is shown in lowercase letters; boxes are drawn around the conserved regions of the aligned pair of sequences.	39
Figure 2.3	A partial comparison matrix between the unknown sequence (lowercase letters listed along the left side) and the reference sequence (uppercase letter listed across the top).	39
Figure 2.4	Mapping the pathway through the matrix (a) the final path through the comparison matrix, and (b) the final sequence alignment.	40
Figure 2.5	Geometry definition for the search loops command	42
Figure 2.6	First-order minimization algorithm (a) steepest descents and (b) conjugate gradient methods.	43

## LIST OF FIGURES (cont.)

	<b>Pages</b>
Figure 2.7	Basic steps on MD simulation study. 47
Figure 2.8	The periodic boundary condition in two dimensions. 49
Figure 2.9	Schematic of the structure-based drug design process 51
Figure 3.1	(a) NMR solved solution structures of the N-terminal domain (residues 1 – 55) and (b) their regularized mean structure (residues 1 – 47) 55
Figure 3.2	Differing in the nature of the metal coordination by the two histidine residues (His12 and His16), designated as D and E forms. 55
Figure 3.3	Crystal structure of the core only domain in (a) a dimeric form and (b) a trimeric form. The structure was colored by specified chain. 56
Figure 3.4	(a) Superimposition of 15 core domain crystal structures, both orientations of Asp116 are highlighted in blue and green, respectively, as for non-cacodylate and cacodylate forms, (b) the orientation of the three acidic residues in the active site region and a $Mg^{2+}$ , and (c) the correspondence RMSD values for 15 core structures. 57
Figure 3.5	(a) Set of structures resolved from NMR technique of the C-terminal domain and (b) the mean structure. 58



## LIST OF FIGURES (cont.)

	<b>Pages</b>
Figure 3.6	59
Crystal structure of the CORE-N (PDB code 1K6Y, <b>51</b> ) (a) in a tetramer form and (b) chain A of the CORE-N was selected in our study, the coordination of Zn <sup>2+</sup> and the two histidine and two cysteine residues in the N-terminal domain were shown in ball and stick model.	
Figure 3.7	60
Crystal structure of the CORE-C (PDB code 1EX4, <b>53</b> ) (a) in a dimer form, CHAPS heteroatoms were shown in ball and stick model, chain A of the structure was selected in this study (b).	
Figure 3.8	66
Superimposition of the four N-terminal regions, (a) only mutant form shown (red), (b) D (violet) and E (cyan) form, note that His12 are colored blue while His16 are colored orange, (c) inclusion of the N-terminal region from the CORE-N (black), and (d) their correspondence RMSD values. The two histidine residues of the structures were shown in ball and stick model.	
Figure 3.9	68
(a) Superimposition of the core region of the CORE (black), the CORE-N (red) and the CORE-C (blue) and (b) the RMSD values.	
Figure 3.10	69
(a) Superimposed structure between the C-terminal domain (CTER, yellow) and the C-terminal region in the CORE-C (green) and (b) the correspondence RMSD values.	
Figure 3.11	70
A homology module in the Insight II program, the main window shows the molecule and the amino acid sequence window demonstrates the homologous sequences	
Figure 3.12	72
The CORE-N, CORE-C and the superimposed structures.	

## LIST OF FIGURES (cont.)

	<b>Pages</b>
Figure 3.13 The full-length model proposed in this study. Each domain was labeled.	73
Figure 3.14 Basic information flows for module in AMBER7.	74
Figure 3.15 The FULL+ION in a solvate box.	76
Figure 3.16 The Hex main window and the docking control panel	80
Figure 4.1 Structure of the CORE-N, two missing regions (residues 47 – 55 and 140 – 148) are zoom out and shown in ball and stick model.	82
Figure 4.2 Structure of the CORE-C, missing region (residues 142 – 145) is shown in ball and stick model and zoom out.	83
Figure 4.3 Amino acid sequence of the full-length HIV-1 IN	84
Figure 4.4 (a) Schematic representation of the full-length structure and (b) superimposition of the two available proposed full-length models	85
Figure 4.5 Full-length structure proposed in this study	85
Figure 4.6 Dimer model of the full-length HIV-1 IN, chain A (red) and chain B (blue).	86
Figure 4.7 The orientation of the catalytic residues (Asp64, Asp116 and Glu152) in the core domain region of both chain A (red) and chain B (blue). The three acidic residues were shown in ball and stick model and labeled.	87

## LIST OF FIGURES (cont.)

	<b>Pages</b>
Figure 4.8 Tetrameric form of the full-length HIV-1 IN, chain A (red), chain B (blue), chain C (green) and chain D (yellow)	87
Figure 4.9 Superimposition between the core only domain, CORE and various core structures from the following structures; CORE-N (black line), CORE-C (red line), FULL (green line) and FULL+ION (blue line).	89
Figure 4.10 Superimposition of the various N-terminal regions in CORE-N (black line), FULL (red line) and FULL+ION (green line) onto the NTER only domain.	89
Figure 4.11 Superimposition of the various C-terminal regions in CORE-C (black line), FULL (red line) and FULL+ION (green line) onto the CTER only domain.	91
Figure 4.12 Superimposition of the two-domain region, the core connected to the N-terminal domain, CORE-N with those regions in the FULL (black line) and FULL+ION (red line). The two high-RMSD regions were shown in circle.	93
Figure 4.13 Superimposition of the two-domain region, the core connected to the C-terminal domain, CORE-C with those regions in the FULL (black line) and FULL+ION (red line).	94
Figure 4.14 Superimposition of the full-length structure without (FULL) and with ions (FULL+ION).	94

## LIST OF FIGURES (cont.)

		<b>Pages</b>
Figure 4.15	Superimposed structures between our FULL+ION model (red) with Luca <i>et al.</i> 's model (green) and their correspond RMSD per residue (b), superimposed structure between our model with Podtelezhnikov <i>et al.</i> 's structure (blue) (c). The different between the core-N linkage in the CORE-N (orange) and Luca <i>et al.</i> 's (green) structure was also displayed (d).	95
Figure 4.16	The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE system.	96
Figure 4.17	Global RMSD value of the CORE system with respect to the average structure over the time range of 2-ns.	97
Figure 4.18	RMSD per residue of the CORE system with respect to the average structure over the time range of 2-ns.	98
Figure 4.19	Stereo views showing the superimposition of the CORE structure taken every 100 ps through 2 ns MD trajectory.	99
Figure 4.20	Distances between the Mg <sup>2+</sup> and carbonyl oxygen atoms of Asp64 and Asp116 (a), torsional angle measurement for Asp64, Asp116 (b) and Glu152 (c).	99
Figure 4.21	Distances between the Mg <sup>2+</sup> and the carbonyl oxygen atoms of the two aspartate residues Asp64 (d1 and d2 as for OD1 and OD2, respectively) and Asp116 (d3 and d4 for OD1 and OD2, respectively) of the CORE system.	100
Figure 4.22	Torsional angles ( $\chi_a$ ) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE system.	101

## LIST OF FIGURES (cont.)

		<b>Pages</b>
Figure 4.23	The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE-N system.	102
Figure 4.24	Global RMSD value of the CORE-N system with respect to the average structure over the time range of 2-ns.	102
Figure 4.25	Distances between the Mg <sup>2+</sup> and the carbonyl oxygen atoms of the two aspartate residues Asp64 ( <i>d1</i> and <i>d2</i> as for OD1 and OD2, respectively) and Asp116 ( <i>d3</i> and <i>d4</i> for OD1 and OD2, respectively) of the CORE-N system.	103
Figure 4.26	RMSD per residue of the CORE-N system with respect to the average structure over the time range of 2-ns.	103
Figure 4.27	Stereo views showing the superimposition of the CORE-N structure taken every 100 ps through 2 ns MD trajectory.	104
Figure 4.28	Torsional angles ( $\chi_a$ ) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE-N system.	104
Figure 4.29	The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the CORE-C system.	106
Figure 4.30	Global RMSD value of the CORE-C system in core region (a) and C-terminal region (b) with respect to the average structure over the time range of 2-ns.	106
Figure 4.31	Global RMSD per residue of the CORE-C system with respect to the average structure over the time range of 2-ns.	107

**LIST OF FIGURES (cont.)**

	<b>Pages</b>
Figure 4.32 Superimposition of the structure taken every 100 ps from 2 ns MD trajectory for the CORE-C.	107
Figure 4.33 Distances between the Mg <sup>2+</sup> and the carbonyl oxygen atoms of the two aspartate residues Asp64 (d1 and d2 as for OD1 and OD2, respectively) and Asp116 (d3 and d4 for OD1 and OD2, respectively) of the CORE-C system.	109
Figure 4.34 Torsional angles ( $\chi_a$ ) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the CORE-C.	109
Figure 4.35 The total (red line), kinetic (green line) and potential (blue line) energies (a), and the temperature (b) over the 2-ns MD simulation for the FULL system.	110
Figure 4.36 RMSD value of the FULL system in the N-terminal domain (black line), core domain (red line), C-terminal domain (green line) and the whole structure N-core-C (blue line) with respect to the average structure over the time range of 2-ns.	111
Figure 4.37 RMSD per residue of the FULL system with respect to the average structure over the time range of 2-ns.	112
Figure 4.38 Superimposition of the structure taken every 100 ps from 2 ns MD trajectory of the FULL.	112
Figure 4.39 Torsional angles ( $\chi_a$ ) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c) of the FULL.	113
Figure 4.40 The total (red), kinetic (green) and potential (blue) energies (a), and the temperature (b) over the 2-ns MD simulation for the FULLION system.	114

## LIST OF FIGURES (cont.)

		Pages
Figure 4.41	Global RMSD value of the FULLION system in the N-terminal domain (black line), core domain (red line), C-terminal domain (green line) and the whole structure N-core-C (blue line) with respect to the average structure over the time range of 2-ns.	114
Figure 4.42	RMSD per residue of the FULLION system with respect to the average structure over the time range of 2-ns.	115
Figure 4.43	Superimposition of the structure taken every 100 ps from 2 ns MD trajectory of the FULL+ION.	115
Figure 4.44	Distances between the Mg <sup>2+</sup> and the carbonyl oxygen atoms of the two aspartate residues Asp64 (d1, d2) and Asp116 (d3, d4) of the FULLION system.	116
Figure 4.45	Torsional angles ( $\chi_a$ ) plot of three catalytic residues, Asp64 (a), Asp116 (b) and Glu152 (c).	116
Figure 4.46	Distance between each catalytic residues, Asp64-Asp116 (black line), Asp64-Glu152 (red line) and Asp116-Glu152 (green line) over the 2 ns simulation time for the FULL (top) and the FULL+ION (bottom).	121
Figure 4.47	Structure of the CORE-C showing where the labeled residues in the C-terminal region (Ser230, Glu246 and Arg262) was reported to bind with DNA (a) and the distribution of the distance between these key residues in the two-domain CORE-C and the FULL+ION.	124
Figure 4.48	RMSD over the trajectories of the core domain region (residues 56 – 209) of CORE (black line), CORE-C (red line), CORE-N (green line) and FULL+ION (blue line).	125

## LIST OF FIGURES (cont.)

		Pages
Figure 4.49	RMSD over the trajectories (a) of the C-terminal domain (residues 210 – 270) of CORE-C (green line) and FULL+ION (black line) and (b) of the N-terminal domain (residues 1 – 55) of CORE-N (red line) and FULL+ION (black line).	125
Figure 4.50	RMSD over the trajectories (a) of the two-end terminal domain (CORE-C, residues 56 – 270) of CORE-C (green line) and FULL+ION (black line) and (b) of the CORE-N (residues 1 – 210) of CORE-N (red line) and FULL+ION (black line).	126
Figure 4.51	Probability distribution of the angle $\alpha$ representing relative orientation of the core, N-terminal and C-terminal domains of our full-length HIV-1 integrase, FULL+ION (thick line) and FULL (thin line). The arrows are the corresponding values calculated from the Podtelezchnikov's and Luca's models.	127
Figure 4.52	Distribution of the distances $P(R_{x-y})$ between each catalytic triad (see text for details); Asp64-Asp116 (a), Asp64-Glu152 (b) and Asp116-Glu152 (c) throughout the 2-ns simulations for the FULL+ION (solid line) and the FULL (dash line). Single value extracted from Podtelezchnikov's (dash arrow) and Luca's (solid arrow) models were also displayed.	128
Figure 4.53	Distribution of the distance $P(d)$ between $Mg^{2+}$ ion and the carboxyl oxygen of the catalytic triad residues, $d1$ , $d2$ , $d3$ and $d4$ , as for Asp64:OD1, Asp64:OD2, Asp116:OD1, Asp116:OD2, respectively. The arrows illustrate the data extracted from Luca's structure.	130



## LIST OF FIGURES (cont.)

	<b>Pages</b>
Figure 4.54	132
Distribution of the torsional angles ( $\chi_a$ ) of each catalytic triad; $\chi_{64}$ (a), $\chi_{116}$ (b) and $\chi_{152}$ (c), for the FULL+ION (solid line) and the FULL (dash line) systems. The values measured from Podtelezhnikov's (dash arrow) and Luca's (solid arrow) models were also plotted.	
Figure 4.55	133
Dimer full-length complexed with viral DNA in side view (a) and front view (b). Subunits A and B of the dimer were colored as red and blue, respectively.	
Figure 4.56	134
A wild type tetramer HIV-1 IN complexed with both viral and host DNA proposed in this study, top view (a) and side view (b). Each monomer unit was colored as red, blue, yellow and green for chains A, B, C, and D, respectively.	
Figure 4.57	135
A model of wild type tetrameric HIV-1 IN complexed with DNA. The integrase monomers are colored as red, blue, yellow and blue as for chains A, B, C and D, respectively.	

## NOTATION

		<b>Section</b>
<b>CHAPTER 1</b>		
AIDS	Acquired Immune Deficiency Syndrome	1.1.1
HIV-1	Human Immunodeficiency Virus type 1	1.1.1
MD	Molecular Dynamics	1.1.3
CDC	The Centers for Disease Control and Prevention	1.2.1
UNAIDS	The Jointed United Nations Programme on HIV/AIDS	1.2.2
WHO	World Health Organization	1.2.2
CD4	A large glycoprotein molecule on the surface of T lymphocyte.	1.2.3
CTL	Cytotoxic T-lymphocyte	1.2.3
NEUT AB	Neutralizing Antibody	1.2.3
HIV-2	Human Immunodeficiency Virus type 2	1.3.1
GAG	Group specific Antigen gene	1.3.2
POL	Polymerase	1.3.2
ENV	Envelope	1.3.2
MA	Matrix	1.3.2
CA	Capsid	1.3.2
NC	Nucleocapsid	1.3.2
RT	Reverse Transcriptase	1.3.2
PR	Protease	1.3.2
IN	Integrase	1.3.2
LTR	Long Terminal Repeat	1.3.2
SU	Surface	1.3.2

**NOTATION (cont.)**

		<b>Section</b>
TM	Transmembrane	1.3.2
RTI	Reverse Transcriptase Inhibitor	1.4.1
PI	Protease Inhibitor	1.4.1
FDA	The Food and Drug Administrator	1.4.1
NRTI	Nucleoside/Nucleotide Reverse Transcriptase Inhibitor	1.4.1
nNRTI	Non-nucleoside Reverse Transcriptase Inhibitor	1.4.1
QSAR	Quantitative Structure Activity Relationship	1.4.2
COMFA	Comparative Molecular Field Analysis	1.4.2
COMSIA	Comparative Molecular Similarity Index Analysis	1.4.2
NMR	Nuclear Magnetic Resonance	1.5.2
HTH	Helix Turn Helix	1.5.2.1
SH3	Src Homology 3	1.5.2.3
ATT	Attachment sites	1.5.4
DNA	Deoxyribonucleic acid	1.6.1
RNA	Ribonucleic acid	1.6.1
A	Adenine	1.6.2
T	Thymine	1.6.2
C	Cytosine	1.6.2
G	Guanine	1.6.2
U	Uracil	1.6.2

## NOTATION (cont.)

		Section
<b>CHAPTER 2</b>		
SCR	Structurally Conserved Region	2.1.1
AMBER	Assisted Model Building with Energy Refinement	2.2.1
PME	Particle Mesh Ewald	2.2.5
FFT	Fast Fourier Transform	2.2.5
 <b>CHAPTER 3</b>		
CORE-N	Core domain connected to N-terminal domain (1K6Y)	3.1.1
CORE-C	Core domain connected to C-terminal domain (1EX4)	3.1.1
PDB	Protein Data Bank	3.1.1
RMSD	Root Mean Square Deviation	3.1.1.1.2
CORE	Core domain (1BL3)	3.1.1.1.2
CHAPS	3-[(3-chloramidopropyl)dimethylammonia]-1-propanesulfonate	3.1.1.2.2
CAC	cacodylate ion	3.1.1.2.2
CAS	s-(dimethylarsenic)-cysteine	3.1.1.2.2
CAF	s-dimethylarsinoyl-cysteine	3.1.1.2.2
TTA	tetraphenyl-arsonium	3.1.1.2.2
TTO	(3,4-dihydroxy-phenyl)-triphenyl-arsonium	3.1.1.2.2
FULL	Full-length HIV-1 IN	3.1.3.1
FULL+ION	Full-length HIV-1 IN with ion	3.1.3.1

**NOTATION (cont.)**

		<b>Section</b>
<b>CHAPTER 4</b>		
NTER	N-terminal domain (1WJC)	4.1.1
CTER	C-terminal domain (1IHV)	4.1.1