



# **THESIS**

**MOLECULAR DYNAMICS AND COMBINED QM/MM  
MODELLING OF HIV-1/RT ACTIVE SITE: DISCRIMINATING  
BETWEEN ALTERNATIVE MECHANISMS IN DNA  
POLYMERIZATION**

**THANYADA RUNGROTMONGKOL**

**GRADUATE SCHOOL, KASETSART UNIVERSITY**

**2006**



**THESIS APPROVAL**  
**GRADUATE SCHOOL, KASETSART UNIVERSITY**

\_\_\_\_\_  
Doctor of Philosophy (Physical Chemistry)

**DEGREE**

\_\_\_\_\_  
Physical Chemistry

**FIELD**

\_\_\_\_\_  
Chemistry

**DEPARTMENT**

**TITLE:** Molecular Dynamics and Combined QM/MM Modelling of HIV-1 RT  
Active Site: Discriminating between Alternative Mechanisms in  
DNA Polymerization

**NAME:** Miss Thanyada Rungrotmongkol

**THIS THESIS HAS BEEN ACCEPTED BY**

\_\_\_\_\_  
**THESIS ADVISOR**

( Associate Professor Supa Hannongbua, Dr.rer.nat. )

\_\_\_\_\_  
**COMMITTEE MEMBER**

( Mr. Adrian J. Mulholland, Ph.D. )

\_\_\_\_\_  
**COMMITTEE MEMBER**

( Associate Professor Supanna Techasakul, Ph.D. )

\_\_\_\_\_  
**DEPARTMENT HEAD**

( Associate Professor Noojaree Prasitpan, Ph.D. )

**APPROVED BY THE GRADUATE SCHOOL ON** \_\_\_\_\_

\_\_\_\_\_  
**DEAN**

( Associate Professor Vinai Artkongharn, M.A. )

# **THESIS**

## **MOLECULAR DYNAMICS AND COMBINED QM/MM MODELLING OF HIV-1/RT ACTIVE SITE: DISCRIMINATING BETWEEN ALTERNATIVE MECHANISMS IN DNA POLYMERIZATION**

**THANYADA RUNGROTMONGKOL**

**A Thesis Submitted in Partial Fulfillment of  
the Requirements for the Degree of  
Doctor of Philosophy (Physical Chemistry)  
Graduate School, Kasetsart University  
2006**

**ISBN 974-16-2665-7**

Thanyada Rungrotmongkol 2006: Molecular Dynamics and Combined QM/MM  
Modelling of HIV-1 RT Active Site: Discriminating between Alternative  
Mechanisms in DNA Polymerization. Doctor of Philosophy (Physical Chemistry),  
Major Field: Physical Chemistry, Department of Chemistry. Thesis Advisor:  
Associate Professor Supa Hannongbua, Dr.rer.nat. 139 pages.  
ISBN 974-16-2665-7

We have investigated the structure and dynamics of the HIV-1 RT active site, by modelling the active conformation of the HIV-1 RT/DNA/dTTP ternary complex. This has included molecular dynamics simulations and combined QM/MM modelling. Three different systems were studied to investigate the effects of different protonation states of dTTP (a deprotonated and two different mono-protonated triphosphate forms), and the effects of different possible protonation state of potentially catalytic aspartate residues (Asp185 and Asp186) were tested. The model of the deprotonated form of dTTP (model A) with the two aspartates in their charged (basic) form seemed to be the most stable and its orientation was in good agreement with crystal structure. The main aim is to study and investigate the details of reaction mechanism in DNA polymerization. We have proposed three reaction steps; Step 1 is the deprotonation of the terminal 3'-OH group of the primer stand; Step2 is the nucleophilic attack of the negatively charged terminal 3'-OH group on P $\alpha$  atom of dTTP; and Step 3 is the P $\alpha$ -O3 $\alpha$  breaking down to gain the final product complex. Three different base mechanisms (Asp185, Asp186, and dTTP) for H-transfer reaction following by nucleotide addition were estimated with two different semiempirical (AM1 and PM3) QM/MM methods. The most feasible H-transfer reaction was found to proceed in a corresponding reaction path via Asp185 which plays an important role as the catalytic base. Consequently, the nucleophilic attacking on P $\alpha$  of dTTP (Step 2) leading to the formation of the pentacovalent intermediate and the subsequent P $\alpha$ -O3 $\alpha$  breaking bond of this structure. Step 3 generates the creation of the 3'-5' newly formed phosphodiester and pyrophosphate resulting in the elongation of the primer stand by one new nucleotide and the leaving group, respectively. The activation barrier for overall reaction is energetically closed to 18.4 kcal/mol in which the rate-limiting step is the H-transfer reaction to Asp185 (model A). The critical structures along the reaction pathway were stabilized by the H-bond interactions with Lys65, Arg72 and Asp113 and some tightly bound water molecules. These studies highlight the utility of QM/MM molecular dynamics simulations for calculation of free energy profiles for enzyme reactions. The results provide insight into the structure and interactions of the active site of this important enzyme, with implications for its mechanism, which may be useful in inhibitor design.

---

Student's signature

---

Thesis Advisor's signature

/ /

## TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	i
LIST OF TABLES	iii
LIST OF FIGURES	iv
ABBREVIATIONS	x
INTRODUCTION	1
Significant and background of the problem	1
HIV-1 RT as a molecular target for drug therapy	4
Anti-AIDS drug therapy and drug-resistance	8
Objectives	12
LITERATURE REVIEW	13
Combined QM/MM modelling of chemical reactions in large systems	13
Applications to enzymatic reactions	15
Structure of the HIV-1 RT/dsDNA/dTTP ternary complex in respect with other DNA polymerase	22
METHODS OF CALCULATIONS	27
Structure of the HIV-1 RT/dsDNA/dTTP ternary complex	27
Preparation of simulation system	28
Molecular dynamics simulations	29
QM/MM simulation procedure	32
RESULTS AND DISCUSSION	36
Molecular dynamics investigation on the RT/dsDNA/dTTP ternary complex	36
Molecular dynamics simulations	36
Coordination of the two magnesium ions, and the dNTP binding site	45
Protonation states of the aspartic sidechain	51

**TABLE OF CONTENTS (cont'd)**

	Page
Combined quantum mechanical and molecular mechanical approach on the RT/ds/DNA/dTTP ternary complex	56
QM/MM minimizations	56
QM/MM dynamics simulations	61
QM/MM approach in conjunction with umbrella sampling method	64
Mechanistic investigation in the small QM/MM partition scheme	64
Mechanistic investigation in the large QM/MM partition scheme	68
Deprotonated triphosphate system (Model A)	69
Mono-protonated triphosphate system (Model B)	80
CONCLUSIONS	90
LITERATURE CITED	93
APPENDIX	109
Appendix A: Theoretical background	110
Appendix B: Publications, Proceedings and Presentations	121

## LIST OF TABLES

Table		Page
1	Force constants values for buffer restraints ( $\text{kcal mol}^{-1} \text{\AA}^{-2}$ )	30
2	Average RMSD of protein backbone atoms and all protein heavy atoms in the polymerase active sites from the two 3 500 ps simulations: (1) Model A (deprotonated dTTP); (2) Model B (mono-protonated dTTP on the $\text{O}\gamma 2$ -oxygen) with respect to the crystal structure of the HIV-1 RT/DNA/dTTP ternary complex	41
3	Selected structural parameters in Models A and B averaged over 3 500 ps MD simulations for each model	41
4	Hydrogen bond frequencies for protein– and water–phosphate hydrogen bonds averaged over 3 500 ps MD simulations in Models A and B	48
5	Atom distances ( $\text{\AA}$ ) averaged over 500 ps MD simulations in four different active-site models with either D185 or D186 protonated for both Models A and B: (1) HD185-Model A; (2) HD185-Model B; (3) HD186-Model A; (4) HD186-Model B	52
6	RMSD of the active-site amino acids (D110-Y115 and M184-D186) from the four 500 ps of simulations: (1) HD185-Model A; (2) HD185-Model B; (3) HD186-Model A; (4) HD186-Model B with respect to the crystal structure	55
7	Atom distances of minimized structures from MM (CHARMM27) and QM/MM (AM1/CHARMM27 and PM3/CHARMM27) energy minimizations in Models A and B (from the final MD structure) compared to the crystal structure of HIV-RT [36] with its missing 3'-OH primer terminus	58
8	The average structural parameters between $\alpha$ -phosphorus and oxygen or carbon in the nucleophile addition	67
9	Some parameters of critical structures in mechanistic reaction of HIV-1 RT via D185 as normal base	72

## LIST OF FIGURES

Figure		Page
1	Overview of HIV-1 replicative cycle	3
2	Structure of the HIV-1 RT catalytic complex. A view of the RT catalytic complex with the polymerase active site on the left and the RNase H domain on the right. The domains of p66 are in color: fingers (red), palm (yellow), thumb (orange), connection (cyan), and RNase H (blue); p51 is in gray. In the two chains, the domains have very different relative orientations. The DNA template strand (light green) contains 25 nucleotides, and the primer strand (dark green), 21 nucleotides. The dNTP is in gold.	6
3	Representation of the HIV-1 RT active site obtained from the 1RTD crystal structure.	7
4	Structures of selected non-nucleoside reverse transcriptase inhibitors (NNRTIs)	9
5	Structures of the anti-HIV nucleoside analogues (NRTIs)	10
6	Possible reaction pathways for ATP hydrolysis. Three different mechanisms of water activation are shown, each with its proton transfer steps in a different color: the violet arrow corresponds to the direct path, red arrows correspond to the Ser181 path, and green arrows correspond to the Ser236 path. The dashed arrow shows the attack of the activated water on the $\gamma$ -phosphate. All the atoms treated with QM in the calculations are shown explicitly.	16
7	Two phosphoryl transfer schemes provided by theoretical calculations. Scheme A is the concerted phosphoryl and proton-transfer model suggested by previous semiempirical calculations and some DFT calculations; scheme B described a dissociative phosphoryl transfer and the shift of the proton to Asp166, which was suggested by the most recent DFT calculations.	19



# LIST OF FIGURES (cont'd)

Figure		Page
8	A general pathway for nucleotide incorporation by DNA polymerases. E, dNTP, and PPi refer to DNA polymerase, 2'-deoxyribonucleoside 5'-triphosphate, and pyrophosphate, respectively. DNA <sub>n</sub> and DNA <sub>n+1</sub> represent the DNA before and after the nucleotide incorporation to the 3'-primer terminus. Red and green distinguish the crystal closed and open states of poly/DNA complexes. Pink and yellow arcs represent pol β's conformational closing and opening (before and after the chemical reaction) motions. (c) Active-site coordination of the nucleotide-binding (A) and catalytic (B) Mg <sup>2+</sup> in the closed ternary pol β/DNA/dCTP complex. In part c, the <sub>A</sub> Mg <sup>2+</sup> coordinates the α-, β-, and γ-phosphate oxygens, D190, D192, and a water molecule; and, the <sub>B</sub> Mg <sup>2+</sup> coordinates the α-phosphate oxygen, primer O3', D190, D192, D256, and a water.	24
9	Two modelled structure represented for the active site: (a) the small QM/MM partition scheme; (b) the large QM/MM partition scheme. QM region is shown in style of bond and stick.	34
10	RMSDs (from the initial structure) of all heavy atoms as a function of simulation time, for two systems studied here: (1) Model A (deprotonated dTTP); (2) Model B (mono-protonated dTTP on the Oγ2-oxygen). Three 700 ps MD simulations are shown in the plot with different shades.	37
11	Root mean square fluctuations (RMSF) (Å) of protein backbone atoms averaged over 3 500 ps MD simulations from two systems shown in different shades: (1) Model A (deprotonated dTTP) in black; (2) Model B (mono-protonated dTTP on the Oγ2-oxygen) in white. RMSF for all residues is shown in 11(a) and for some selected residues in 11(b).	38

# LIST OF FIGURES (cont'd)

Figure		Page
12	Root mean square fluctuations (RMSF) (Å) of all heavy atoms by nucleotide averaged over 3 500 ps MD simulations for the DNA template [12(a)], DNA primer and dTTP substrate (S) [12(b)] from the two simulations shown in different shades: (1) Model A in black; (2) Model B in white. The template base for substrate (ST) and primer terminus (PT) are Ade5 and Cyt6 while the primer terminus (P) is Gua22.	40
13	(a) Definition of angle $\theta$ and distance d. Plot of angle $\theta$ as a function of time in Models A (b) and B (c). Plot of distance as a function of time in Models A (d) and B (e). Three 700 ps MD simulations are shown in the plot with different shades.	43
14	Conformation of mono-protonated triphosphate on O $\gamma$ 3 oxygen of dTTP substrate, two Mg <sup>2+</sup> ions and their coordination environments (Model C) in the final structure from MM MD simulation. Hydrogen atoms are not shown except for the mono-protonated O3 $\gamma$ atom and the bound water molecules. The colours were used by atom type colour.	47
15	Superimposition of minimized CHARMM at last MD snapshot (Model A systems: blue-violet, Model B systems: red-violet) compared to the crystal structure (blue): 15(a) HD185-Model A; 15(b). HD185-Model B; 15(c) HD186-Model A; 15(d). HD186-Model B. Hydrogen atoms are not shown except for the mono-protonated O2 $\gamma$ atom and the bound water molecules.	53

# **LIST OF FIGURES (cont'd)**

Figure		Page
16	Superimposition of QM/MM minimized AM1/CHARMM (orange) and PM3/CHARMM (yellow) structures (from the final MD structure) compared to the crystal structure (blue). Model A is shown in 16(a) and Model B in 16(b). Hydrogen atoms are not shown except for the mono-protonated O2 $\gamma$ atom (Model B) and the bound water molecules.	60
17	Plot of selected distances and angles as a function of time from two different semiempirical QM/MM dynamics simulations for 50 ps: (a) AM1/CHARMM on Model A; (b) PM3/CHARMM on Model A; (c) AM1/CHARMM on Model B; (d) PM3/CHARMM on Model B.	62
18	Free energy profiles for reaction coordinate as hydrogen primer terminus-carboxyl oxygen (OD1) of Asp185.	64
19	Modeled structure represented for the active site. QM region is shown shaded with four hydrogen link atoms cycled.	66
20	Pentacovalent intermediate, magnesium coordination environment and possible hydrogen bonds taken from AM1 QM/MM simulation snapshot. H atoms bonded to the C and N atoms are not shown.	67
21	Free energy profiles for three different H-transfer reactions in Model A from H3T-hydrogen on the 3'-OH primer terminus to a H-acceptor (Asp185, Asp186 or dTTP) using reaction coordinate as different reaction coordinate between breaking and forming bond.	71
22	Reactant complex (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	72

# **LIST OF FIGURES (cont'd)**

Figure		Page
23	Product complex (Model A) from H-transfer reaction via Asp185, magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	73
24	Free energy profiles for nucleophilic addition in reaction coordinate as: different reaction coordinate between $P\alpha-O3\alpha$ breaking and $P\alpha-O3'$ forming bond.	77
25	Pentacovalent intermediate (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	78
26	Final product complex (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	79
27	Free energy profiles for three different H-transfer reactions in Model B from H3T-hydrogen on the 3'-OH primer terminus to a H-accepter (Asp185, Asp186 or dTTP) using reaction coordinate as different reaction coordinate between breaking and forming bond.	82
28	Reactant complex (Model B), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	83

### LIST OF FIGURES (cont'd)

Figure		Page
29	Product complex (Model B) from H-transfer reaction via Asp185, magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules and on triphosphate moiety are shown.	84
30	Free energy profiles for nucleophilic addition in reaction coordinate as: different reaction coordinate between $P\alpha-O3\alpha$ breaking and $P\alpha-O3'$ forming bond.	86
31	Pentacovalent intermediate (Model B), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	87
32	Final product complex (Model B), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.	88
Appendix Figure		Page
A1	Sequence of program step required to solve the Roothaan-Hall equations, Self Consistent Field procedure	116

## ABBREVIATION

ABNR	=	Adopted basis Newton-Raphson
Ade	=	Adenine
AIDS	=	Acquired Immune Deficiency Syndrome
Ala (A)	=	Alanine
AM1	=	Austin Model 1
AOs	=	Atomic Orbitals
Asn (N)	=	Asparagine
Arg (R)	=	Arginine
Asp (D)	=	Aspartic acid
		Beck's three parameter hybrid functional using the LYP
B3LYP	=	correlation functional
CD4	=	Cluster of differentiation 4
Cyt	=	Cytosine
Cys (C)	=	Cysteine
DFT	=	Density Functional Theory
DNA	=	Deoxyribonucleic acid
dNTP	=	Deoxynucleoside triphosphate
dsDNA	=	Double-stranded deoxyribonucleic acid
dTTP	=	Deoxythymidine triphosphate
Gln (Q)	=	Glutamine
Glu (E)	=	Glutamic acid
Gly (G)	=	Glycine
Gua	=	Guanine
H-transfer	=	Proton transfer
HF	=	Hartree-Fock theory
His (H)	=	Histidine
HIV-1	=	Human Immunodeficiency Virus Type 1
Ile (I)	=	Isoleucine
Leu (L)	=	Leucine

LCAO	=	Linear combination of atomic orbitals
LCAO-MO	=	Linear combination of atomic orbitals to molecular Orbitals
Lys (K)	=	Lysine
MD	=	Molecular Dynamics
Met (M)	=	Methionine
MLR	=	Multiple linear regression
MM	=	Molecular Mechanics
MNDO	=	Modified neglect of diatomic overlap
MO	=	Molecular Orbitals
MOs	=	Molecular Orbital
MP2	=	Second order Möller-Plesset
mRNA	=	Messenger ribonucleic acid
NDO	=	Neglect of diatomic differential overlap
NNRTIs	=	Non-Nucleoside Reverse Transcriptase Inhibitors
NRTIs	=	Nucleoside Reverse Transcriptase Inhibitors
Phe (F)	=	Phenylalanine
PM3	=	Modified neglect of diatomic overlap, parametric method number 3
PPi	=	Pyrophosphate
Pro (P)	=	Proline amino acid
QM	=	Quantum Mechanics
QM/MM	=	Quantum Mechanical/Molecular Mechanical method
RMS	=	Root Mean Square
RMSD	=	Root Mean Square Deviation
RNA	=	Ribonucleic acid
RNaseH	=	Ribonuclease H
RT	=	Reverse Transcriptase
SBMD	=	Stochastic Boundary Molecular Dynamics method
SD	=	Steepest descent
Ser (S)	=	Serine
SCF	=	Self-consistent field

Thr (T)	=	Threonine
tRNA	=	Transfer ribonucleic acid
Trp (W)	=	Tryptophan
Tyr (Y)	=	Tyrosine
Thy	=	Thymine
UNAIDS	=	The Joint United Nations Programme on HIV/AIDS
Val (V)	=	Valine
WHO	=	World Health Organization