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

## INTRODUCTION

Significant and background of the problem

'Acquired Immune Deficiency Syndrome' (AIDS) is one of the most important diseases of human over twenty years now. More than twenty-five million people have died of AIDS since 1981 and millions more in developing and transitional countries in the world continue become infected with HIV (Human Immunodeficiency Virus). By the end of 2005, an estimated 38.6 million people worldwide were living with HIV/AIDS and around three million deaths from AIDS, despite recent improvements in access to antiretroviral treatment reported by UNAIDS/WHO.

As same as all types of virus, HIV cannot grow by its own but must infect the living cells to make its new copies. HIV is in a special class called retroviruses and within the subgroup of lentiviruses. The T-cell or CD4 cell is the host of HIV. HIV seeks out the CD4 cells and then proteins on the HIV surface attach to complimentary proteins on the CD4 cell as a 'lock and key' type system. HIV injects its proteins into cytoplasm which causes the cell membrane fusion to the outer envelope of HIV. The protective coating surrounding viral RNA is dissolved to release its genetic material (RNA) for reproduction. Consequently, the single stranded viral RNA is converted to the double stranded DNA by HIV-1 reverse transcriptase (HIV-1 RT). RT uses building blocks from the T-cell to change the single stranded viral RNA to DNA which contains the genetic information to the mode of genetic storage in the host cell needed for the HIV reproduction. The genetic information from the reverse

transcription process is compatible with human genetic material. This newly formed DNA is integrated into the cell nucleus, where it is spliced into the human DNA by the HIV-1 integrase. HIV must wait for more protein building blocks to complete the reproductive process. When the cell becomes activated, the HIV genes are treated like human genes. First it converts them into messenger RNA (mRNA) using human enzymes. Then mRNA is transported outside the nucleus, and is used as a blueprint for producing new HIV proteins and enzymes. Among the mRNA strands produced by the cell are complete copies of HIV genetic material. These gather together with newly made HIV proteins and enzymes to form new viral particles which are ready to release from the cell. HIV-1 protease plays a vital role at this stage of the HIV life cycle by chopping up long strands of protein into smaller pieces, which are used to construct mature viral cores. The newly matured HIV particles are ready to infect another cell and begin the replication process all over again as shown in Figure 1.
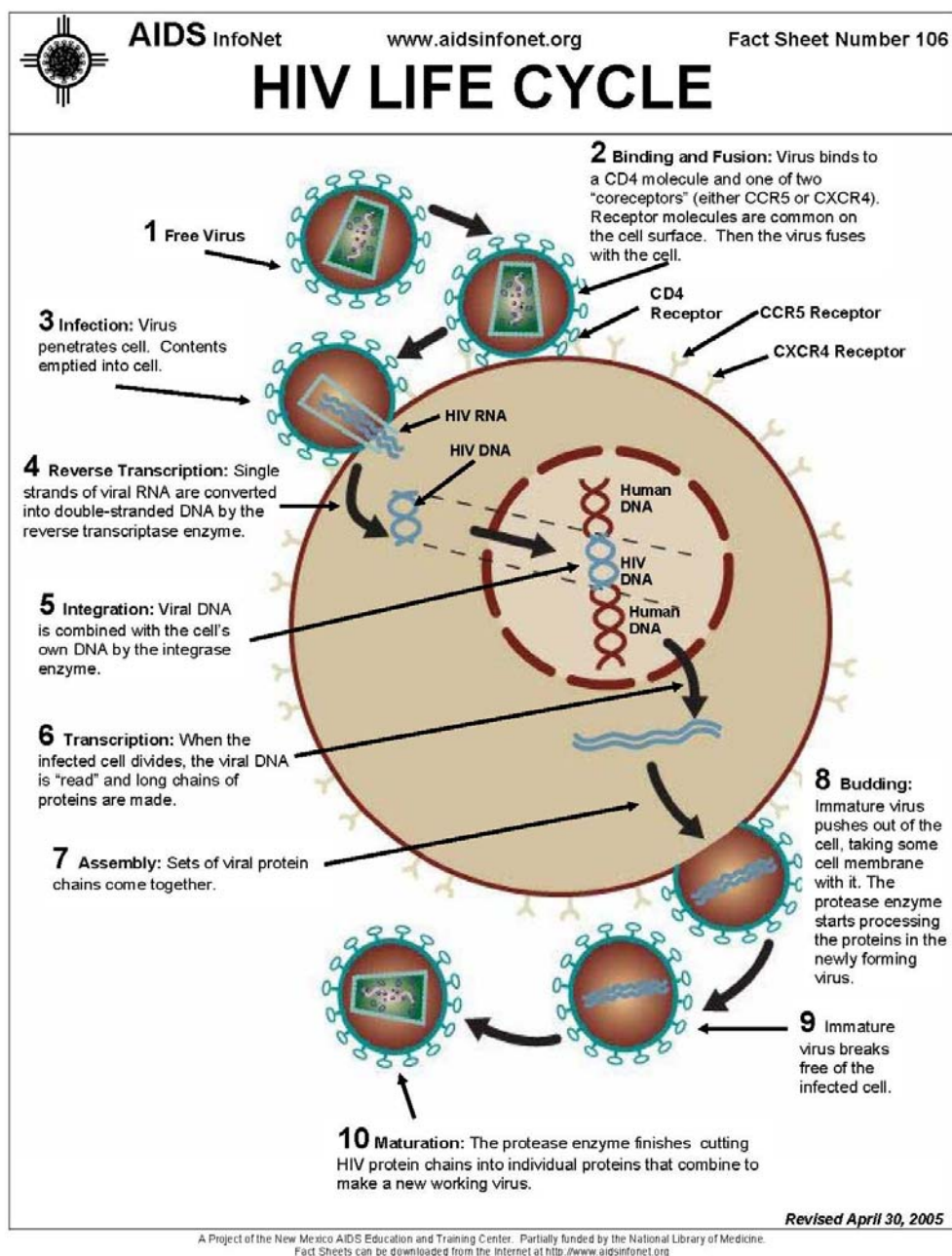
Figure 1   Overview of HIV-1 replicative cycle

Source: Aidsinfonet (2005)

HIV-1 RT as a Molecular Target for drug therapy

HIV-1 reverse transcriptase (RT) has received a great deal of attention within the past several years due to the importance of its role in the life cycle of the HIV-1 virus. RT, a multifunctional enzyme of the human immunodeficiency virus (HIV), catalyses the multi-step conversion of the single-stranded viral RNA genome into a double-stranded DNA copy which can then be integrated into the host genome (Wondrak *et al.*, 1986). During catalysis of the multi-step conversion, RT has to bind successively the following different primer/templates: tRNA/RNA, DNA/RNA, RNA/DNA and DNA/DNA. Due to its key role in the HIV life cycle, RT is an important target for antiviral agents in the treatment of acquired immunodeficiency syndrome (AIDS). RT is a heterodimer composed of p66 and p51 subunits. Both subunits contain 'finger', 'palm', 'thumb', named according to their resemblance to a right hand, and connection subdomains, with the ribonuclease H (RNaseH) domain found only in the p66 subunit (Hostomsky *et al.*, 1992). The p51 subunit is originated from p66 by proteolytic cleavage of the C-terminal RNaseH domain. Consequently, both subunits have the same amino acid sequence. However, the structural organization of the subdomains is completely different, resulting in an asymmetric dimer with the active site only located in the p66 subunit. RT exhibits both DNA polymerase and RNaseH activities to complete the reverse transcription. The p66 subunit containing the catalytic activity region has a functional polymerase active site and a large cleft that binds the template-primer DNA, a DNA-binding cleft, formed by the p66 fingers, palm and thumb subdomains. Even though only the p66 subunit seems to play an important role in the catalytic function of HIV-1 RT, the p51 subunit is necessary for the enzyme to be fully active conformation of the p66 subunit. For this reason, the p66 subunit would be focused on discussing the catalytic activities of this enzyme. The p66 palm subdomain not only contains the polymerase active site but also the non-nucleoside binding site (allosteric site) buried within RT near the polymerase active site.

There are numerous crystal structures of HIV-1 RT in the Protein Data Bank (PDB) which include those of unliganded RT, RT bound to several NNRTIs and RT complexed to double-stranded (dsDNA) with and without a deoxynucleoside triphosphate (dNTP). Reported crystal structures of RT show the polymerase active site located on the palm subdomain. In unliganded RT, the p66 thumb subdomain is folded into the DNA-binding cleft and lies over the palm subdomain, nearly touching the fingers subdomain, in a "thumb down" configuration. Consequently, the DNA-binding cleft is closed. In contrast, the nonnucleoside binding pocket only exists in the structures of RT complexed to a NNRTI, its formation being probably induced by proximity of the inhibitor. The ternary complex of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase with a DNA template:primer and a deoxynucleoside triphosphate (dNTP), HIV-1 RT/dsDNA/dTTP, has been solved at 3.2 Å resolution (PDB entry code 1RTD.PDB) (Huang *et al.*, 1998) as shown in Figure 2. The presence of a dideoxynucleotide at the 3'-primer terminus allows capture of a state in which the 3'-hydroxyl group of the terminal primer is poised for attack on the α-phosphorus of dTTP. In contrast, the structure of unliganded HIV-1 RT probably represents another defined stage in the normal catalytic pathway, but only during initiation or reinitiation and not during processive synthesis. In RT/dsDNA/Fab complex (Jacobo-Molina *et al.*, 1993), the thumb is in an upright position, its tip approximately 32 Å away from the fingers. It is likely to be close to an "open" state of the enzyme after translocation to the next register but before binding of the dNTP. Bending of the fingertips from open to closed must occur in synchrony with successive steps in the overall mechanism, as also suggested for enzymes of the Pol I family (Sawaya *et al.*, 1997).

Structural studies of the crystallographic structure comparison and site-directed spin label experiment suggested that RT has several extremely flexible regions which have been proposed to be essential for the polymerization process. Molecular dynamics simulation of unliganded RT and RT complexed to double-stranded DNA (dsDNA) have shown that the RT flexibility depends on its ligation state, increasing upon DNA binding. The motions of amino acids at the non-nucleoside binding pocket upon binding of the NNRTI are anticorrelated to the p66

fingers in the RT/DNA complex and correlated to the RNaseH subdomain in unliganded RT.
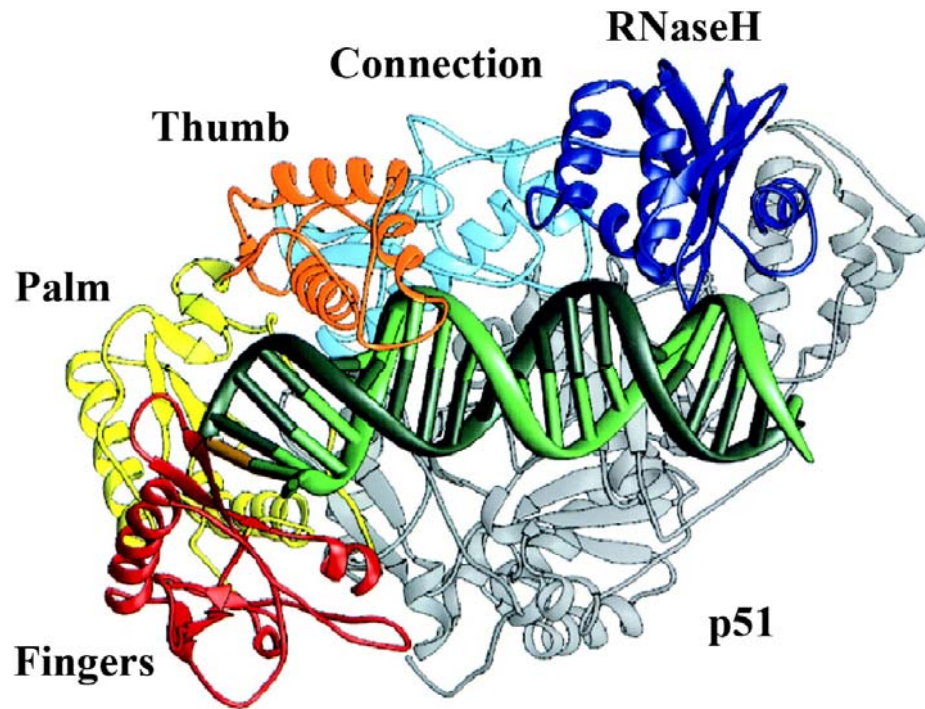
<u>**Figure 2**</u>   Structure of the HIV-1 RT catalytic complex. A view of the RT catalytic complex with the polymerase active site on the left and the RNase H domain on the right. The domains of p66 are in color: fingers (red), palm (yellow), thumb (orange), connection (cyan), and RNase H (blue); 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.

Source: Huang (1998)
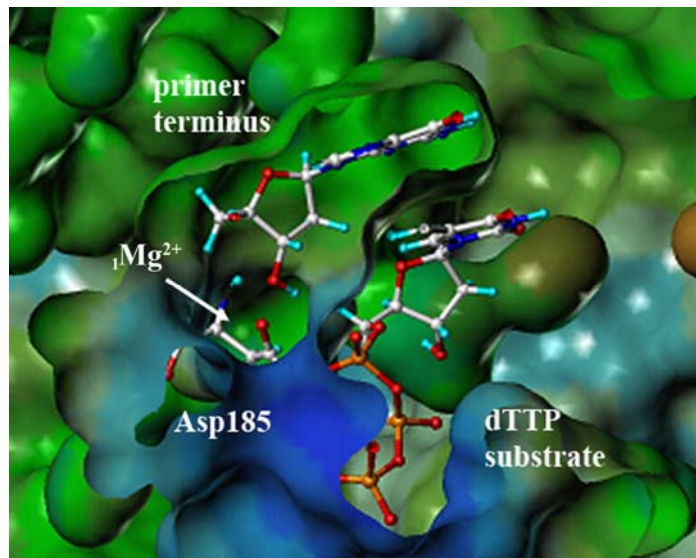
Figure 3   Representation of the HIV-1 RT active site obtained from the 1RTD crystal structure.

Source: Rungrotmongkol (2004)

The polymerase activity of HIV-1 RT shares several features (e.g. the metal chelation and the binding of deoxyribonucleoside triphosphate (dNTP) in polymerase active site) with other DNA polymerases such as the Klenow fragment, T7 DNA polymerase and RB69 pol α polymerase (Steitz. 1999; Doublie *et al.*, 1998; Beese *et al.*, 1993; Wang *et al.*, 1997; Bibillo *et al.*, 2005). DNA polymerization in polymerases is thought to involve a two metal ion mechanism (Beese *et al.*, 1991; Derbyshire *et al.*, 1991), with these being magnesium ions in the case of HIV-1 RT. The catalytic magnesium ion ($_1Mg^{2+}$) interacts with the 3'-hydroxyl group of the primer terminus and decreases its pK_a value to facilitate attack on the α-phosphate of deoxythymidine triphosphate (dTTP) (see Figure 3). RT incorporates deoxyribonucleoside triphosphate (dNTP) by elongation at the 3'-hydroxyl group of the primer terminus, forming 3'–5' phosphodiester bonds, and releasing pyrophosphate. This leads to the elongation of the DNA primer by one new nucleotide. This essential step in the retroviral life cycle is targeted by a variety of drugs used to combat AIDS.

Anti-AIDS drug therapy and drug-resistance

A significant and growing problem in anti-AIDS drug therapy is drug resistance caused by mutation of the RT enzyme. Anti-HIV-1 RT drugs can be divided into two main classes: non-nucleoside reverse transcriptase inhibitors (NNRTIs) and nucleoside reverse transcriptase inhibitors (NRTIs) which have been found to be strong inhibitors of this enzyme and are in clinical.

NNRTIs, e.g. nevirapine (Dipyridodiazepinones), TIBO (tetrahydroikmidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one and thione), HEPT (1-[(2-hydroxyethoxy)-methyl]-6-(phenylthio)thymine) and Efavirenz ((-)-6-chloro-4-cyclopropylethynyl-4-trifluoromethyl-1,4-dihydro-2h-3,1-benzoxazin-2-one) (Hannongbua *et al.*, 2001; Saen-Oon *et al.*, 2003), are non-competitive inhibitors which are highly specific for HIV-1 RT at a common allosteric site approximately 10 Å from the polymerase active site. The structures of NNRTIs are given in Figure 4. It is thought that these act by distorting the HIV-1 RT enzyme in several important regions near the polymerase active site (e.g. a displacement of the catalytic aspartate residues), forcing it to adopt an inactive conformation (Lawtrakul *et al.*, 2004). NNRTIs also affect the regions that contact the nucleic acid (Hannongbua *et al.*, 2001; Saen-Oon *et al.*, 2003). The rapidity of the selection of drug resistant HIV in patients was such that single-point mutations in the virus made first generation NNRTIs such as nevirapine unusable in mono-therapy. Among the mutations in RT that were orginally described for nevirapine resistance were those found at Tyr181 and Tyr188, both of which gave rise to high level resistance. Mutation at Tyr181 has since been frequently reported in resistance studies for many other NNRTIs and the change is almost always to cysteine. In the case of the codon 188 mutation, a greater variety of changes are reported, nevirapine and HEPT select the Y188C mutation, whereas Y188H or Y188L are selected with TIBO or α-APA.

8-Cl *TIBO* : tivirapine   nevirapine   delavirdine

HEPT   loviride

First generation NNRTIs

efavirenz   emivirine
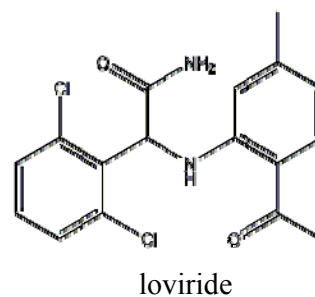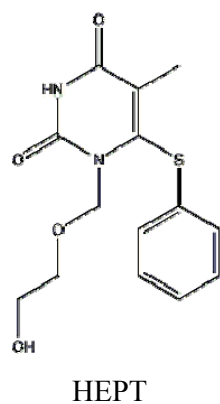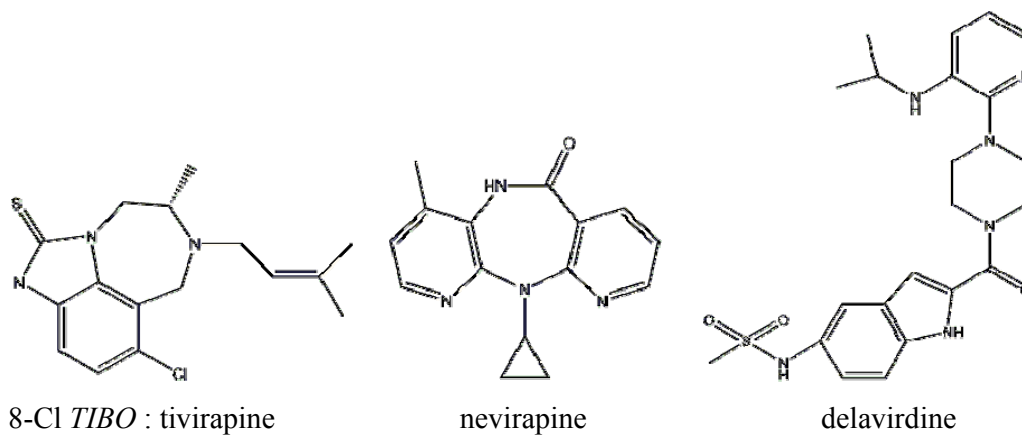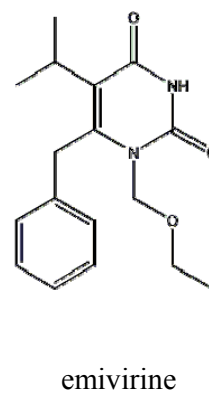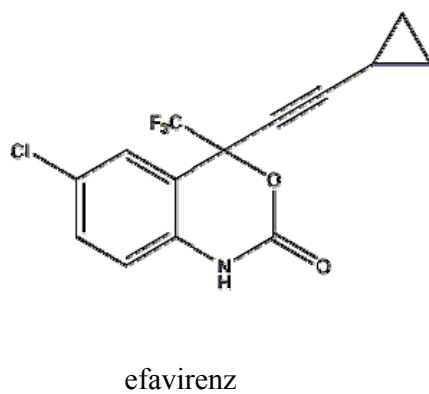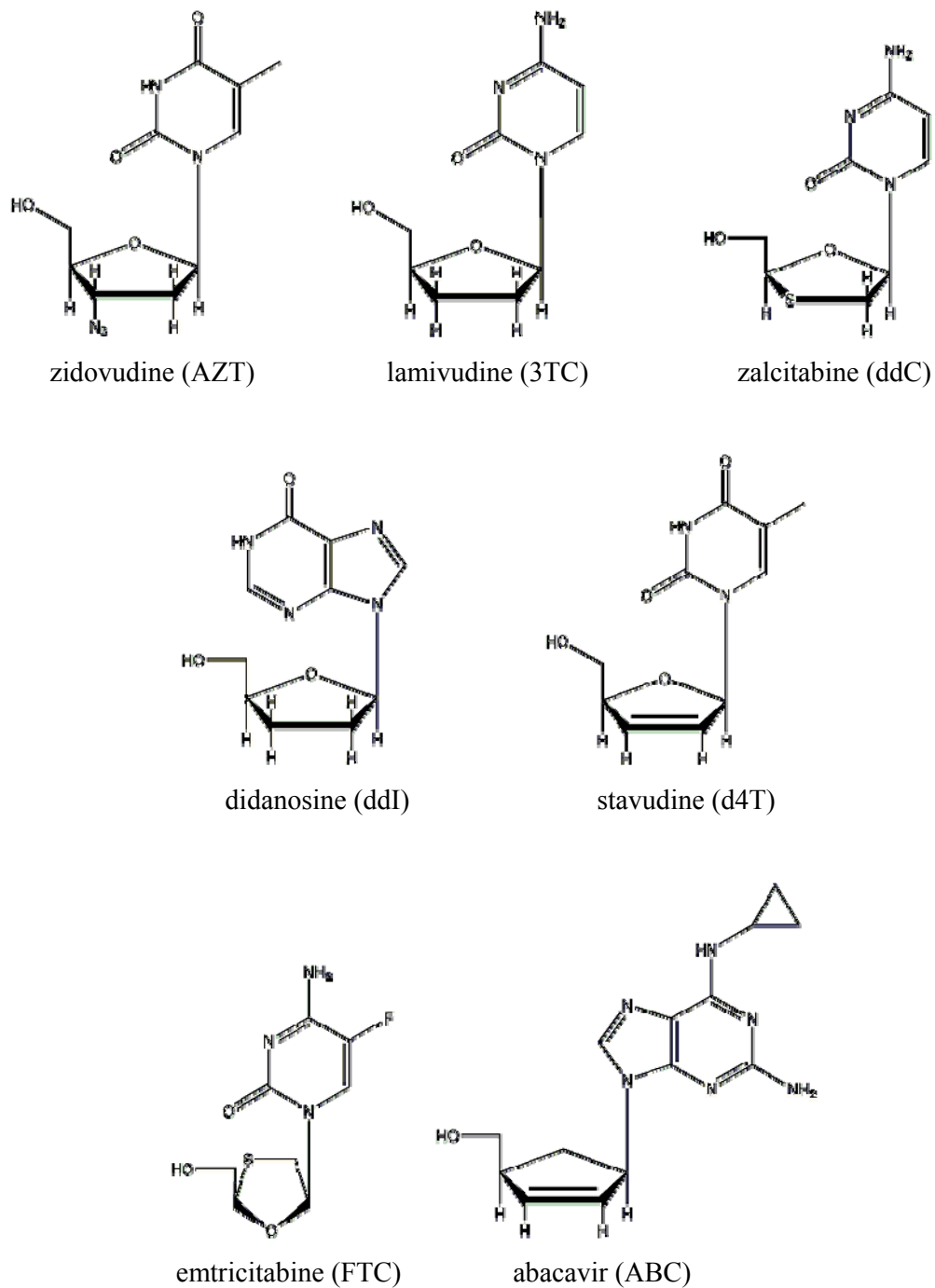
Second generation NNRTIs

Figure 4   Structures of selected non-nucleoside reverse transcriptase inhibitors

(NNRTIs)

Source: Hannongbua (2006)

zidovudine (AZT)   lamivudine (3TC)   zalcitabine (ddC)

didanosine (ddI)   stavudine (d4T)

emtricitabine (FTC)   abacavir (ABC)

Figure 5   Structures of the anti-HIV nucleoside analogues (NRTIs)

NRTIs, such as AZT(3'-azidodeoxythymidine), ddC (dideoxycytidine) and ddI (dideoxyxymosine) (Chong *et al.*, 2002; Kedar *et al.*, 1990) shown in Figure 5, are competitive inhibitors of the nucleotide substrate which lack the 3'-OH group at the terminal DNA primer found on normal nucleosides and act as chain terminators when incorporated into viral DNA by HIV-1 RT. Therefore, these can engage in the first chemical step of the polymerization reaction but then block DNA elongation at the polymerase active site of the host cell. Important mutations giving rise to NRTI resistance include M41L, A62V, K65R, D67N, K70R, V75I, F77L, F116Y, Q151M, M184V/L, T215Y/F and K219Q. Biochemical studies suggest that two main factors are involved in HIV mutation to resist NRTIs: change in drug/substrate molecular recognition (Klarmann *et al.*, 2000; Gao *et al.*, 2000; Sluis-Cremer *et al.*, 2000) such as resistance to the 3TC analogues (for example, the M184V mutant has increased steric hindrance, blocking the appropriate binding of 3TC derivatives but still permitting the incorporation of normal nucleosides) and mutations (K65R) giving cross-resistance to ddI, ddC and 3TC; and rescue of DNA synthesis by increased repolymerization reaction (Meyer *et al.*, 1998) such as AZT resistance associated with M41L, D67N, K70R, T215Y/F, K219E/Q.

This growing problem of drug resistance makes finding new potent inhibitors a vital goal: novel inhibitors should be potentially useful as pharmaceutical lead compounds in the development of new drugs. In spite of intensive experimental investigations, the detailed origins of the binding affinity of inhibitors interacting in polymerase active site of HIV-1 RT remain unclear. Also, understanding of the mechanism of polymerization may assist in the development of more effective NRTIs. The enzyme may bind inhibitors strongly if they resemble substrates or, perhaps even better, unstable species in the mechanism such as intermediates and transition states. Better understanding of the reaction mechanism of RT could help with the design of new inhibitors.

Objectives

In this present study, combined QM/MM and molecular dynamics methods have been applied to the active site of HIV-1 RT bound to DNA primer, deoxythymidine triphosphate (dTTP) and two magnesium ions with the aims of:

1. Studying the structural and dynamical properties of two different protonation states of the dTTP substrate with either Asp185 or Asp186 (or neither) protonated in the HIV-1 RT active site.

2. Investigating the mechanistic reactions of deprotonation of the 3'-OH terminal primer via three different H-accepters (Asp185, Asp186 and dTTP substrate), followed by nucleophilic attack to the α-phosphorus of dTTP and then releasing pyrophosphate.

LITERATURE REVIEW

**Combined QM/MM modelling of chemical reactions in large systems**

It is noted nowadays that the best computational technique in studying the structural aspects of matter and a diversity of its physical and chemical properties as well as chemical reactions involves the use of quantum mechanics. The computational expense associated with high level quantum mechanical calculations imposes a rigorous limitation on the system size that can be studied. In practice, the size of many systems of interest in biology or material sciences makes impossible to be treated with quantum mechanical methodologies. A way to tackle this problem is that only the groups directly involved the reaction are considered whereas the remainder of system and solvent is neglected. Moreover, this method to study such processes must be able to describe chemical reactions within large systems (nano–scale systems, biological systems, supramolecular assembles).

Combined quantum mechanical/molecular mechanical (QM/MM) methods make possible the modelling of chemical reactions in large systems such as enzymes (Ridder *et al.*, 2003b; Mulholland. 2005) which the basic idea was originated (Warshel *et al.*, 1976) for the study of the catalytic mechanism of lysozyme. In QM/MM calculations, the system of interest is divided into two regions; the small region treated quantum mechanically (any level of *ab initio* QM (semiempirical, HF, MPn, CI, CC, MCSCF) or DFT) and the reminder of system considered classically by computationally less expensive molecular mechanics. The quantum region includes all atom directly involved the chemical reaction (i.e., a small region such as the substrate and selected amino acids in the active site) and the rest of the system (the surrounding protein and solvent), believed to change little during chemical process, is treated with a molecular mechanics force field. Therefore, the combined QM/MM method is becoming increasingly popular and powerful for investigating the structure, properties and chemical reactions of large systems, specially the biological systems.

QM/MM methods have been used to study many enzyme-catalysed reactions (Ranaghan *et al.*, 2004; Hermann *et al.*, 2005; Gregersen *et al.*, 2003). They can provide detailed mechanistic information which is difficult or impossible to obtain from experiments alone. However, QM/MM calculations (particularly at high levels of QM theory) (Hermann *et al.*, 2005) can be very demanding of computational resources because of the large size of enzymes. Many biochemical QM/MM applications have used semiempirical molecular orbital (MO) methods such as AM1 or PM3 (Van Der Vaart *et al.*, 2000; Khandogin *et al.*, 2004; Ridder *et al.*, 2002; Sheppard *et al.*, 2000) to reduce the computational time. However, the semiempirical methods have a number of limitations concerning their accuracy and reliability. For example the AM1 Hamiltonian is deficient in the treatment of H–bond and the AM1 treatment on phosphorus shows a spurious and very sharp potential barrier at 3.0Å distorting the conformational geometries and giving spurious activation barriers, while both AM1 and PM3 grossly underestimates the dispersion interaction. It is possible to carry out QM/MM calculations with *ab initio* (Sheppard *et al.*, 2000; Lyne *et al.*, 1999) or density-functional theory (DFT) (Sheppard *et al.*, 2000; Lyne *et al.*, 1999; Stanton *et al.*, 1995) QM approaches, but these are more expensive when the QM subsystem has many atoms, and prohibitively expensive when molecular simulations are required.

Applications to enzymatic reactions

Over last decade, the large amount of publications shows that the combined QM/MM techniques are acceptable as a powerful tool to studying chemical reactions in large systems, especially enzyme. Amongst the many applications of the QM/MM method are studies on the mechanistic reactions catalyzed by acetyl cholinesterase (Fuxreiter *et al.*, 1998), aldose reductase (Varnai *et al.*, 2000), carbonic anhydrase (Toba *et al.*, 2000; Merz Jr. *et al.*, 1997; Lu *et al.*, 1998), catechole *O*–methyltransferase (Lau *et al.*, 1998; Zheng *et al.*, 1997; Kuhn *et al.*, 2000), chorismate mutase (Kharyin *et al.*, 1999 ; Marti *et al.*, 2001), citrate synthase (Mulholland *et al.*, 1998; Mulholland *et al.*, 2000), cylcophilin (Li *et al.*, 2003), cytochrome c oxidase (Olsson *et al.*, 2005), cytochrome P450 (Bathelt *et al.*, 2003; Bathelt *et al.*, 2004; Bathelt *et al.*, 2005), dihydroxyfolate reductase (Cummins *et al.*, 2000), 4-chlorobenzoyl-CoA dehalogenase (Xu *et al.*, 2004), enolase (Liu *et al.*, 2000), formate dehydrogenase (Torres *et al.*, 1999), glycoxalase I (Feierberg *et al.*, 2000), the GTPase reaction of p21 RAS (Glennon *et al.*, 2000), glutathione-S-transferase (Ridder *et al.*, 2002), haloalkane dehydrogenase (Lau *et al.*, 2000), HhaI methyltransferase (Lau *et al.*, 1999), 4-hydroxyphenylpyruvate dioxygenase (Borowski *et al.*, 2004), *p*-hydroxybenzoate hydroxylase (Ridder *et al.*, 2003a), Class A β-lactamase (Hermann *et al.*, 2005), a metallo β-lactamase (Park *et al.*, 2005); orotidine 5'–monophosphate decarboxylase (Warshel *et al.*, 2000), naphthalene dioxygenase (Bassan *et al.*, 2004), neutrophil elastase (Gleeson *et al.*, 2004), phenol hydroxylase (Ridder *et al.*, 2000), protein tyrosine phosphatase (Kolmodin *et al.*, 1999b ; Kolmodin *et al.*, 1999a), ribonuclease A (Glennon *et al.*, 1998), class III ribonucleotide reductase (Cho *et al.*, 2004), thermolysin (Antonczak *et al.*, 1998), trypsin (Kollman *et al.*, 2001), 4-oxalocrotonate tautomerase (Cisneros *et al.*, 2004); triosephosphate isomerase, brovine protein tyrosine phosphatase, uracil DNA-glysosylase (Dinner *et al.*, 2001). Some of the most recently published work was presented with the intention of illustrating the QM/MM techniques which applied on modelled metalloenzymatic systems, in this thesis, involved the magnesium ion and phosphate molecule.
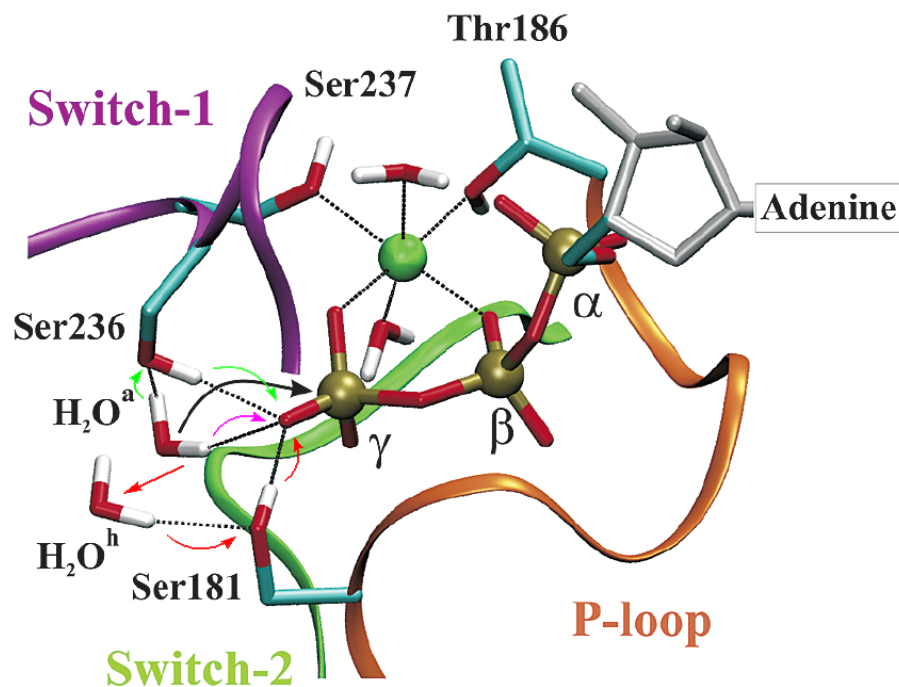
Figure 6   Possible reaction pathways for ATP hydrolysis. Three different mechanisms of water activation are shown, each with its proton transfer steps in a different color: the violet arrow corresponds to the direct path, red arrows correspond to the Ser181 path, and green arrows correspond to the Ser236 path. The dashed arrow shows the attack of the activated water on the γ-phosphate. All the atoms treated with QM in the calculations are shown explicitly.

Source: Schwarzl (2006)

The myosin motor drives a variety of essential motility processes. Myosin II is comprised of a heavy chain and two light chains. The heavy chain contains the catalytic site and the actin-binding site at the N-terminal domain, consequently, only the N-terminal domain able to hydrolyzed ATP and move along an actin filament. In myosin, the ATP binding pocket is formed by three loops: these are the phosphate-binding loop (P-loop), the Swith-1 loop and the Swith-2 loop. The catalysis of ATP hydrolysis takes place only when myosin is in the Swith-1 closed/Swith-2 closed conformation. At the active conformation, the water molecules can move in and out the closed active site. The solvent-phosphate exchange was observed at three terminal γ-phosphate oxygen atoms. The energy barrier of hydrolysis in myosin is 14.5 kcal/mol which much lower than in solution (c.a. 29 kcal/mol). In addition, the hydrolysis reaction is likely to be a single step mechanism with an unstable intermediate. Two different mechanisms of the phosphate hydrolysis are proposed: these are purely associative and dissociative mechanisms. In the purely associative mechanism, the attacking nucleophile enters to the reaction and then a pentacovalent oxyphosphorane structure is formed as intermediate. In contrast, ATP dissociates into ADP and PO_3^- intermediate prior to the nucleophile enters to the reaction found in the purely dissociative mechanism. There are some experimental evidences (for example, the character of a concerted S_N2 mechanism and the binding of water to the γ-phosphate prior to formation of the transition state) indicating that the hydrolysis proceeds via an associative mechanism rather than a dissociative mechanism. At least, six different associative mechanisms have been proposed, differing in how the attacking water becomes activated.

Three proposed water activation mechanism of a phosphate hydrolysis reaction, namely, the direct, Ser236, and Ser181 mechanisms were explored (Schwarzl *et al.*, 2006) (Figure 6). Combined *ab initio* QM/MM techniques with conjugated peak refinement (CPR) method have been recently used to study the catalytic mechanism of ATP hydrolysis in the myosin motor. The QM region (57 atoms) included the triphosphate moiety of ATP, the Mg^{2+} ion, the two crystal water molecules and all active site residues directly involved in the reaction. During the reaction path calculations, all atoms within 12 Å sphere of γ-phosphorus atom in the

QM region, P-loop, Swith-1 loop and Swith-2 loop were allowed to move freely. Geometry optimization was performed using Hartree Fock calculations for the QM atoms with a 3-21G(d) basis set and subsequently a 6-31G(d,p) basis set. Then the optimized structures with HF/6-31G(d,p) method were single point calculated with density functional theory (DFT) at the B3LYP/6-31+G(d,p) level of theory. The results showed that the lack of electron correlation effects leads to an error in the energy barrier of hydrolysis reaction (c.a. 10 kcal/mol) even though large basis set are used. Although, the density functional theory is higher accuracy, B3LYP still gives an average error (c.a. 3 kcal/mol) and maximum error (c.a. 20 kcal/mol).

The energy barriers of hydrolysis reaction of at any level of theory of three mechanisms are very close to each other indicating that three mechanisms are within equally error. The optimized geometries with HF/6-31G(d,p) method are more reasonable structural accuracy than others. Mostly, the ATP triphosphate geometry from the optimized reactant in myosin is similar to that of the Mg·MTP structure in solution. The results showed the hydrogen bond network (involved Ser181, Ser236, Ser237 and the attacking and helper water molecules) around the γ-phosphate corresponds to the three water activation mechanisms. The resulting reaction pathways, found in the catalytically competent closed/closed conformation of the Switch-1/Switch-2 loops of myosin, are all associative with a pentavalent bipyramidal oxyphosphorane transition state but can vary in the activation mechanism of the attacking water molecule and in the way the hydrogens are transferred between the heavy atoms. In the product state, the coordination bond between the Mg^{2+} metal cofactor and Ser237 in the Switch-1 loop is broken; thereby facilitating the opening of the Switch-1 loop after hydrolysis is completed. This reveals a key element of the chemomechanical coupling that underlies the motor cycle, namely, the modulation of actin unbinding or binding in response to the ATP or ADP·Pi state of nucleotide-bound myosin.
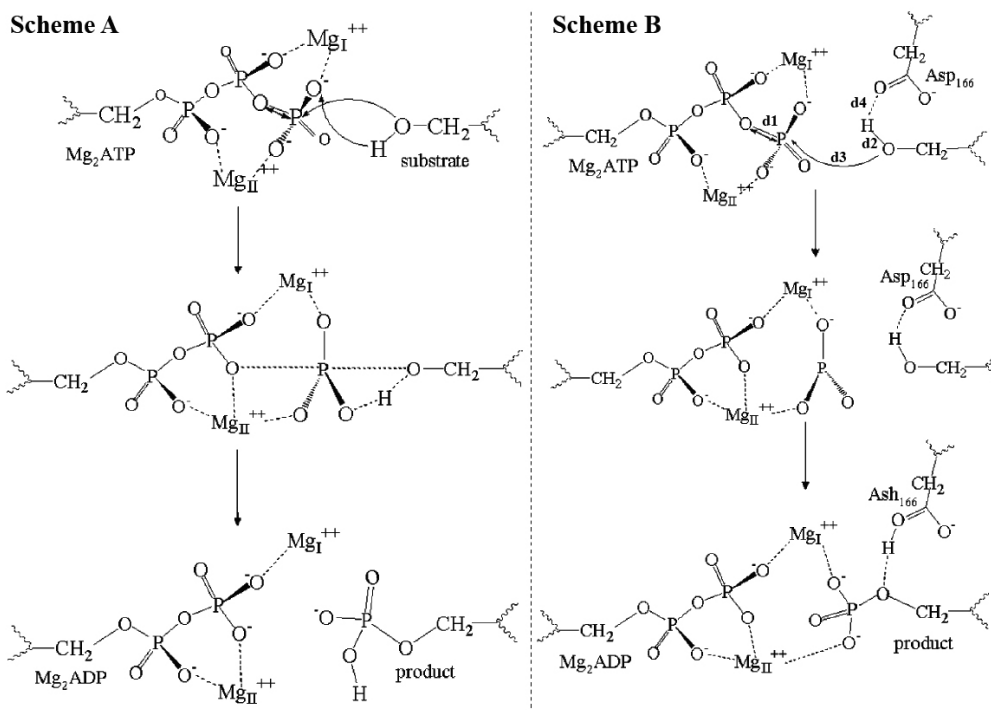
Figure 7   Two phosphoryl transfer schemes provided by previous theoretical calculations. Scheme A is the concerted phosphoryl and proton-transfer model suggested by previous semiempirical calculations and some DFT calculations; scheme B described a dissociative phosphoryl transfer and the shift of the proton to Asp166, which was suggested by the most recent DFT calculations.

Source: Cheng (2005)

Protein phosphorylation, a common reversible covalent modification catalyzed by protein kinases, plays a critical role in cellular regulation and signal transduction. Its function is to catalyze the transfer of the ATP γ-phosphate to the P-site serine. It was some experimental evidences that Asp166 is possible to act as H-acceptor (the catalytic residue) from the hydroxyl group of the substrate Ser which help to increase the nucleophilic reactivity of the P-site serine. However, Asp166 may be the normal base to abstract the hydroxyl hydrogen of the P-site serine after the necleophilic attack reaction due to its pKa (less than 5) and pKa of serine closed to 14 in the aqueous solution. Lys168 is the one of important residues in the phosphorylation reaction but no evidence what Lys168 exactly does in the ATP binding site. It may stabilize the γ-phosphate of ATP or transfer one of its protons to the γ-phosphate. Furthermore, the structural studies mentioned that the glycine-rich loop which covers the ATP molecule may play an important role to stabilize the transition state. Although, the experimental studies (such as the X-ray crystallographic structures and mutation evidences) provided some valuable information at the ATP pocket, many questions in mechanistic detail of the phosphorylation reaction are unanswered and unclear. Consequently, the theoretical approach is a powerful tool to access and investigate the chemical reactions in biological systems. Two different reaction mechanisms of phosphorylation (scheme A and B) were proposed: scheme A is the concerted reaction which the transfers of phosphoryl and proton occur in the same time whereas the reaction of the phosphoryl transfer to the P-site serine followed by the proton transfer from the hydroxyl serine to Asp166 proposed as stepwise reaction shown in scheme B (Figure 7). The concerted reaction in scheme A was previously studied by many theoretical groups with semiempirical and density functional theory calculations. They found that the energy barrier for the concerted reaction were quite high (36-42 kcal/mol). In contrast, the energy barrier of the dissociative (stepwise) reaction using DFT calculations was 11.0 and 16.5 kcal/mol investigated by Valiev and Diaz groups, respectively. However, Henkelman et al. found that the reaction barrier is varied depended on the modelled size and initial crystal structures.

Recently, Cheng *et al.*, 2005 have performed the density functional theory QM/MM methods to study the phosphorylation reaction catalyzed by cyclic AMP-dependent protein kinase (PKA) with two different QM sizes and two different initial crystal structures (1L3R and 1ATP). The QM region (49 atoms or 75 atoms) consisted of the triphosphate group of ATP (or whole), side chains of P-site Ser, Asp166 and Lys168, and two Mg^{2+} ions which all possibly involved to the chemical reaction. First, each QM/MM models was optimized by combined B3LYP QM/MM calculations at the 6-31G* level of theory, leading to an optimized structure for the reactant, then employed the reaction coordinate driving method to search for the transition state and product by an iterative restrained minimization on the reaction coordinate: RC = d1 + d3 - d2 - d4 where definitions of d1 to d4 are given in Figure 7. Consequently, the critical structures (reactant, transition state and tetrahedral intermediate), were carried out with single-point higher level (MP2 and B3LYP) QM/MM calculations with a larger 6-31+G* basis set. As the B3LYP(6-31G*) QM/MM calculations in the small 1L3R model, the reaction barrier of scheme A is more than 20 kcal/mol while only 11.6 kcal/mol found in scheme B compared to the experimental data (13.8 kcal/mol). It is thus indicated that the stepwise reaction of scheme B shows more realistic pathway, and good agreement with experiment and also previously theoretical studies (11.0 kcal/mol by Valiev *et al.* and 16.5 kcal/mol by Diaz *et al.*): Asp166 serves as the catalytic base to accept the proton from the P-site serine after the transition state was formed in the dissociative phosphoryl transfer reaction. On the other hand, Lys168, Asn171, Asp184, and the conserved waters bound to Mg^{2+} ions do not directly contribute to lower the energy barrier of the phosphorylation reaction. The Mg^{2+} ions (especially Mg_{II}), glycine rich loop, and Lys72 are found to stabilize the transition state through electrostatic interactions. Molecular dynamics simulations for 12 ns on both wild type and K168A mutated PKA showed that Lys168 helps maintain the ATP and P-site serine in appropriated conformation to ready for the phosphoryl transfer reaction. The QM/MM calculations with different QM/MM partition schemes or different initial structures yield consistent results.

Structure of the HIV-1 RT/dsDNA/dTTP ternary complex in respect with other DNA polymerase

Based on amino acid sequence comparisons and crystal structure analysis, DNA polymerases can be divided into seven families: A, B, C, D, X, Y and RT. Although sequence similarity are completely different in all polymerase, common is a two-metal-ion coordination with three highly conserved carboxylate-containing residues (such as three aspartates, or two aspartates and a glutamate) in the polymerase active site. Amino acid sequence comparisons and structural studies in all polymerases suggest that the nucleotide addition is catalyzed by a two-metal-ion mechanism (usually Mg^{2+}) proposed (Beese *et al.*, 1991 and Derbyshire *et al.*, 1991). Kinetic studies of several high and medium fidelity DNA polymerases have suggested the rate-limiting reaction conformational changes in both before and after the chemical reaction of nucleotide addition. Yang *et al.*, 2004 studied a general pathway of nucleotide incorporation by DNA polymerases with distinct open and closed states; crystallographic closed and open conformation of polymerase β/DNA complex along the pathway as used for illustration. Eight pol β/DNA complexes were separately constructed to perform molecular dynamics simulation: four models before and four models after the chemical reaction of nucleotide incorporation, each case representing an intermediate structure between closed and open states, and each quadruplet involving a nucleotide-binding magnesium ion, catalytic magnesium ion, divalent cations, or no magnesium ion. Studying the effects of the nucleotide-binding and catalytic magnesium ions on pol β's conformational transitions between closed and open states both before and after the nucleotide chemical incorporation help delineate the various roles of these divalent cations on polymerase mechanism and, in particular, associate them with key slow steps that direct nucleotide inserting apart from the chemical reaction itself. They suggested the following sequence in the reaction pathway of polymerase β for correct nucleotide insertion as shown in Figure 8:

(1)   binding of the dNTP/Mg^{2+} association to the polymerase β/DNA in open state

(2)   binding of the catalytic magnesium to the active site

(3)   relatively fast conformational transition from an open to a closed stateinvolving subtle residue motions

(4)   slow, and possibly rate-limiting, assembly of the key amino acid residues, template bases, Mg^{2+} ions and the primer 3'-OH

(5)   slow, and possibly rate-limiting chemical step of the nucleotidyl transfer reaction

(6)   release of the catalytic magnesium

(7)   relatively fast conformational transition from the closed to open complex state again involving subtle residue motions

(8)   release of the product pyrophosphate (PPi)/ Mg^{2+} unit

Steps 1 and 2 may be interchangeable depending on which magnesium ion first binds to the polymerase/DNA active site.
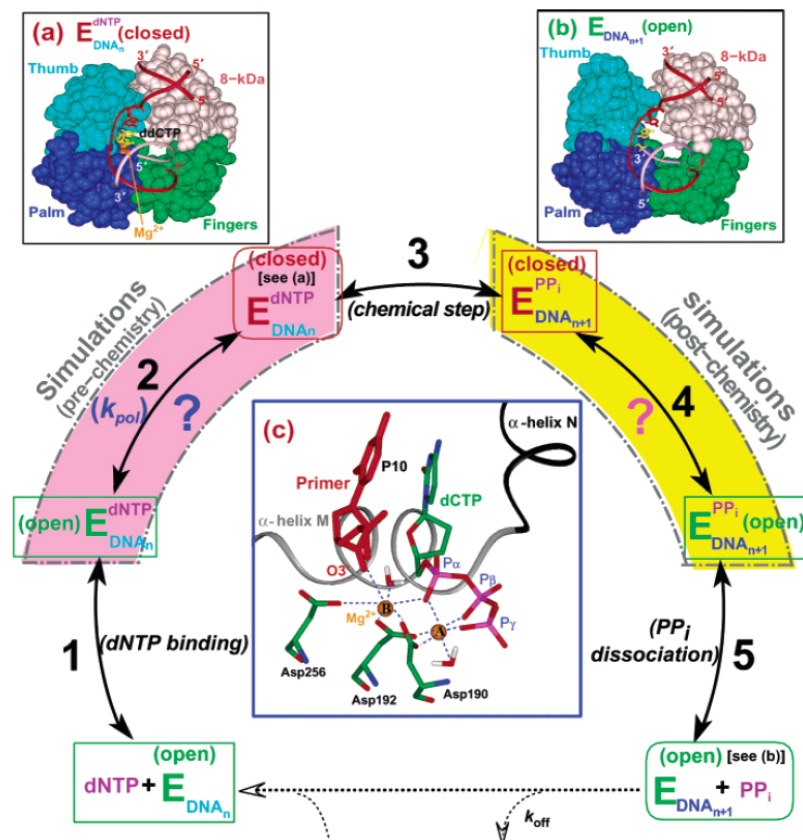
Figure 8   A general pathway for nucleotide incorporation by DNA polymerases. E, dNTP, and PPi refer to DNA polymerase, 2'-deoxyribonucleoside 5'-triphosphate, and pyrophosphate, respectively. DNA_n and DNA_{n+1} represent the DNA before and after the nucleotide incorporation to the 3'-primer terminus. Red and green distinguish the crystal closed and open states of poly/DNA complexes, respectively. Pink and yellow arcs represent pol β's conformational closing (before the chemical reaction) and opening (after the chemical reaction) motions, respectively. (c) Active-site coordination of the nucleotide-binding (label A) and catalytic (label B) Mg^{2+} in the closed ternary pol β/DNA/dCTP complex. In part c, the nucleotidebinding Mg^{2+} coordinates the α-, β-, and γ-phosphate oxygens from dCTP, D190, D192, and a water molecule; and, the catalytic Mg^{2+} coordinates the α-phosphate oxygen from dCTP, primer P10:O3', D190, D192, D256, and a water molecule.

Source: Yang (2004)

Many DNA polymerases are known to incorporate nucleotides to the DNA primer strand through a "two-metal-ion" mechanism in a closed state. In general, the nucleotide-binding ion coordinates the α-, β-, and γ-phosphate oxygens of the incoming dNTP; the catalytic ion coordinates both the α-phosphate of dNTP and the 3′-OH primer. Coordination of the catalytic magnesium ion ($_1Mg^{2+}$) to the 3′-OH primer facilitates the chemical reaction, in-line nucleophilic attack of 3′-OH on the α-phosphorus of dNTP. The position and geometry of the α-phosphorus are stabilized through interactions of the two nonbridging oxygens on the α-phosphorus with protein residues and two functional metal ions in the active site. The nucleotide-binding metal ion assists the departure of the product pyrophosphate from the active site, and both ions help stabilize the structure of the expected pentacovalent transition state in the chemical reaction of nucleotide addition. This "two-metal-ion" functional mechanism has thus been implicated with assembling the catalytic groups, neutralizing the negative charges of the environment, stabilizing the active site structure, and catalyzing the nucleotide insertion.

In the nucleotide binding site of HIV-1 RT (1RTD.PDB) (Huang *et al.*, 1998), the two magnesium ions are coordinated to the dTTP substrate and some amino acids in the active site. The crystal structure lacks the 3'-OH group of the DNA primer terminus which is expected to be a ligand of the catalytic magnesium ion. The other ligands of $_1Mg^{2+}$ are two carboxylate oxygens of D110 (D[Mg1-OD1(D110)] = 2.51 Å and D[Mg1-OD2(D110)] = 2.68 Å), the OD1 carboxylate oxygen of D185 and the α-phosphate of dTTP: these interactions are observed in the crystal structure. In contrast, the nucleotide-binding magnesium ion ($_2Mg^{2+}$) is observed to be octahedrally coordinated, with six ligands: the OD1 carboxylate oxygen of D110 (D[Mg1-OD1(D110)] = 2.13 Å), the backbone carbonyl oxygen of V111, the OD2 carboxylate oxygen of D185 and an oxygen from each phosphate group (α, β and γ) of the dTTP triphosphate. It is noted that only bridging D185 carboxylate group is coordinated symmetrically to the divalent cations whereas the D110 carboxylate group is twisted toward the $_1Mg^{2+}$ ion, with both carboxylate oxygen atoms coordinated to this magnesium ion, and the OD1 carboxylate oxygen acts as a

bridging ligand of two magnesium ions found in the crystal structure. Similar coordination spheres have also been observed in high-resolution structures of ternary complexes of the enzymes DNA polymerase β (Sawaya *et al.*, 1997) and T7 DNA polymerase (Doublie *et al.*, 1998), each of which contain two comparable magnesium ions at the active site. These families of DNA polymerase should be chemically similar in the two-metal ion mechanism of nucleotide addition (catalyzed the phosphoryl transfer reaction) as mentioned in Introduction above. In addition, K65 and R72 in RT are thought to play important roles in stabilizing the dTTP substrate via hydrogen-bonds with its triphosphate moiety. This K65-R72 pair is structurally analogous to the R518-K522 pair and the R754-R758 pair in T7 DNA polymerase and the Klenow fragment, respectively (Kaushik *et al.*, 1996). Other amino acids close to the triphosphate moiety of dTTP are G112, D113, A114, Y115 and Q151 (Sarafianos *et al.*, 1995; Harris *et al.*, 1998).

METHODS OF CALCULATIONS

1. Structure of the HIV-1 RT/dsDNA/dTTP ternary complex

There are numerous crystal structures of HIV-1 RT in the Protein Data Bank (PDB) showing the polymerase active site located on the palm subdomain. Crystal structures of HIV-1 RT with double-stranded DNA have been solved with and without a dNTP substrate. In this study, the starting geometry of the complex was the HIV-1 RT/dsDNA/dTTP ternary complex, whose crystal structure has been solved at 3.2 Å resolution (PDB entry code 1RTD.PDB) (Huang *et al.*, 1998). It is currently the only crystal complex consisting of the reactive part of the enzyme-substrate complex with the double DNA strands, and is therefore the most suitable structure for mechanistic modelling. Protein X-ray crystallography is often unable to resolve the positions of hydrogen atoms, frequently leaving the protonation state of acidic and basic residues, and in this case also of the dTTP triphosphate, unidentified in the crystal structure. In aqueous solution at pH = ~7, the protonation state of the triphosphate moiety can be either fully deprotonated or mono-protonated on the γ-phosphate which has higher pK_a (pK_a = 7.5) than other phosphate groups (Saenger. 1983). In non-metalloenzymes, aspartic acids are most often present in their basic form (i.e. as the carboxylate), though as with other titratable groups in proteins, their pK_as can be significantly altered by their environment (Morikis *et al.*, 2001; Morikis *et al.*, 2003). In RT, the pK_as of Asp185 and Asp186 at the active site may be increased by the proximity two magnesium ions or decreased by electrostatic interactions with neighboring anionic groups (in particular the triphosphate moiety of the dTTP substrate and the two other aspartates, Asp110 and either Asp185 or Asp186). Therefore, the protonation states of these two active-site aspartic acids (Asp185 and Asp186) are unclear. Potentially, either could act as the base (to act as the proton acceptor from the 3'-OH terminal primer), but of course this would not be possible for an aspartic acid which exists in its acidic form. Consequently, the systems with either D185 or D186 (or neither) protonated were tested to see if they could reproduce the basic features of the crystal structure.

2. Preparation of the simulation system

Three different protonation states of the triphosphate moiety were considered: these were fully deprotonated (Model A), or singly protonated on either the Oγ2-oxygen (Model B) or the Oγ3-oxygen (Model C). It was considered unlikely that a proton could be bound to Oγ1-oxygen which is partially coordinated to $_2$Mg^{2+}. To resolve questions concerning the protonation state of two aspartic acids in polymerase active site, we created active-site models with either D185 or D186 protonated for both deprotonated (Model A) and mono-protonated (Model B) forms of the triphosphate moiety. These models are denoted HD185-Model A; HD185-Model B; HD186-Model A; and HD186-Model B. Therefore, in total seven different systems were studied with MM molecular dynamics simulations. Each model was built separately and equilibrated as described previously (Rungrotmongkol *et al.*, 2004). Firstly, the missing 3'-OH of the primer terminus was added with standard (CHARMM) bond lengths and angles to complete the system. Some differences are expected at the active site (e.g. potentially in the coordination of the catalytic magnesium ion ($_1$Mg^{2+})) because the simulated structure is different from the crystal structure by containing the 3'-OH.

RT is a very large enzyme, and to reduce the computing expense, only a truncated section around the active site was modelled, using the stochastic boundary approach. The system is divided into two regions: the reservoir region, which is far away from the polymerase active site, and the reaction zone which contains the active site. The reaction zone is further divided into a reaction region and a buffer region (Brooks *et al.*, 1989). The reservoir region here was eliminated and was compensated for by harmonic forces applied to the buffer region non-solvent heavy atoms, with force constants dependent on the atom type and distance from the reaction centre (Zurek *et al.*, 2004). Water molecules' behavior was influenced by the deformable boundary potential (Brooks *et al.*, 1983b) to maintain them within the simulation sphere. The motion of atoms in the buffer region is governed by Langevin molecular dynamics. The atoms in this region interact with an external heat bath via both random fluctuating and Langevin dissipative forces. Meanwhile, the reaction region

atoms are treated by conventional molecular dynamics via Newton's equations of motion.

3. Molecular dynamics simulations

All simulations were performed with the CHARMM (Brooks *et al.*, 1983a) program, version c27b2. For all simulations with purely MM methods, a cutoff of 13Å was applied to nonbonded interactions and the CHARMM27 parameter set (MacKerell *et al.*, 1998; MacKerell *et al.*, 1995; Foloppe *et al.*, 2000; Mackerell *et al.*, 2000) was used. In the simulations, the system contained a spherical selection of residues centred on the Pα reacting atom in dTTP. All residues and DNA units with at least one atom within 25Å of the centre were included. Hydrogen atoms were added on the basis of CHARMM internal coordinates with the HBUILD facility (Brunger *et al.*, 1988), and were minimized with 2 series of 250 steps of steepest descents (SD) minimization with heavy atoms fixed. Then the system was solvated by superimposing the structure on a cubic box of 8000 pre-equilibrated CHARMM TIP3P water molecules (Jorgensen *et al.*, 1983). All water molecules outside a 26 Å sphere centred on the Pα atom of dTTP were removed and any water molecules of which the oxygen was within 2.8 Å distance of any non-hydrogen atom were deleted. After solvation, the final system consisted of 253 amino acids, 16 DNA units, dTTP substrate, two Mg^{2+} ions and 1056 added water molecules; altogether 7959 atoms with a net charge of -1 for Model A. Models B and C contained 7960 atoms and no net charge (neutral overall). The positions of water molecules were optimized, keeping all other atoms fixed to their initial positions, with 500 steps of steepest descent (SD) and 500 steps of adopted basis Newton-Raphson (ABNR) minimization. A stochastic boundary molecular dynamics simulation was used to equilibrate the water molecules with a buffer radius of 20 Å, while all other atoms were fixed (Brooks *et al.*, 1989). A friction coefficient of $\beta = 62$ ps^{-1} was applied to the water oxygen atoms. The water molecules were restrained to remain in the simulation system (the 25 Å sphere) by a deformable boundary potential (Brooks *et al.*, 1983b). The positions of the water molecules were further optimized with two series of 300 steps of SD and 1500 steps of ABNR minimization, followed by heating from 10 K to 300 K for 1 ps and

equilibration at 300 K for 25 ps (using a 1 fs time step). The positions of water molecules were again optimized with two series of 250 steps of SD and 1500 steps of ABNR minimization.

For stochastic boundary molecular dynamics simulations of the whole system, the boundary of the simulation system was restrained. Each system was divided into two regions based on the starting structure. The reaction region consisted of all atoms within 21 Å and was subjected to no restraints. The buffer region included all atoms lying at distances between 21 Å and 25 Å. The outer, reservoir zone, (containing all atoms further than 25 Å from the centre) was not included in the calculations. Harmonic restraints were applied to non-solvent atoms in the buffer zone, scaled from a maximum at 25 Å to zero at 21 Å from the centre. Force constants for different atom types were calculated from approximate average X-ray temperature factors as given in Table 1 (Zurek *et al.*, 2004). The buffer region was divided into sub-zones (Buffer 1: atoms positioned between 21 and 22 Å, Buffer 2: atoms positioned between 22 and 23 Å, Buffer 3: atoms positioned between 23 and 24 Å and Buffer 4: atoms positioned 24 Å or more from system centre which is α-phosphorus atom).

Table 1 Force constants values for buffer restraints (kcal mol^{-1} Å^{-2})

Atom type / Region	Backbone atoms	Beta carbon atoms	Gamma atoms (carbon, oxygen, sulfer)	Other heavy atoms
Buffer 1	0.27	0.24	0.23	0.22
Buffer 2	0.85	0.78	0.73	0.69
Buffer 3	1.44	1.31	1.22	1.18
Buffer 4	1.70	1.55	1.45	1.40

Charged residues in the buffer region were made artificially neutral (but retaining atomic charges to maintain hydrogen bonds) by scaling the partial charges of the residues in their original geometry to avoid unrealistically large contributions to the electrostatic energy (Zurek *et al.*, 2004; Ranaghan *et al.*, 2003). Therefore, 2 arginine, 6 lysine, an aspartate and 7 glutamate sidechains in this region were neutralized.

The whole structure was minimized with 500 steps of SD and 1250 steps of conjugated gradient (POWELL) minimization, then the system was heated from 10 K to 300 K in three 10 ps intervals: first from 10 K to 100 K, then from 100 K to 200 K, and, finally, from 200 K to 300 K. In the next stage, the whole system was equilibrated for 200 ps. Subsequently, three separate 500 ps simulations were carried out for each of Model A, B and C (i.e. using different initial velocities). 500ps simulations (following 200ps equilibration) were also performed on HD185-Model A, HD185-Model B, HD186-Model A and HD186-Model B. The list of buffer region was updated every 50 steps. Friction coefficients of 250 ps^{-1} for non-hydrogen protein atoms and 62 ps^{-1} for water oxygen atoms were used. SHAKE (Ryckaert *et al.*, 1977; Van Gunsteren *et al.*, 1977) was applied to fix all MM bonds involving hydrogen atoms and a 1 fs time step was used. The structural coordinate from simulations was saved every 50 steps for analysis. Finally, each resulting structure from the last snapshot of simulations was optimized by two series of 250 steps of SD, and then followed by 1500 steps of ABNR minimization.

4. QM/MM simulation procedure

The enzyme-dsDNA-substrate models 1RTD prepared as described above were each partitioned into a QM subsystem and a MM subsystem. Two partition schemes were employed with the sole difference in whether to cut the magnesium coordinating ligands. In the small QM subsystem partition scheme, the QM subsystem comprises the methyltriphosphate of substrate complexed with two Mg^{2+} ions, modeled in its deprotonated form, the methanol portion of deoxyribose ring of the 3'-terminal DNA primer and the carboxylate side chain of Asp185 which is proposed to be a H-acceptor in deprotonation. The small QM region includes 32 atoms including four link atoms with overall charge -1e. The large QM subsystem consists of the methyltriphosphate of the dTTP substrate (modelled in its deprotonated and mono-protonated forms) complexed with two Mg^{2+} ions, the methanol portion of deoxyribose ring of the 3'-terminal DNA primer, the sidechains of Asp110, Asp185 and Asp186, the backbone carbonyl of Val111 terminated with the backbone methyl-amino of Gly112, and a modelled water molecule. Altogether, the large QM region included 58 atoms with an overall charge of -1 in Model A, and 59 atoms with neutral charge in Model B (including 8 link atoms in each case). Both partition schemes use hydrogen 'link atoms' (Field *et al.*, 1990) to model the covalent bonds between carbon atoms across the QM/MM boundary. The link atoms in the large partition scheme are QM hydrogens without classical van der Waals or bonded force field terms. They do, however, interact with all MM charges, where the charges on the link atom host groups have been set to zero (this is the HQ-type link (Reuter *et al.*, 2000) atom which has been found to perform better than the original QQ-type link (Field *et al.*, 1990) atom). In contrast, the link atoms in the small partition scheme are the QQ-type link atoms. Although link atoms are an approximate approach to QM/MM partitioning of covalently bonded systems, they have been found to give good results for a number of simulations of enzyme reactions.

The boundaries of the simulation systems were restrained with conditions similar to the stochastic boundary method of molecular dynamics. Each system was divided into three zones based on the starting structure. The central or "reaction" zone consisted of all atoms within 21 Å and was subject to no restraints. The "buffer" zone included all atoms lying at distances between 21 Å and 25 Å away. Harmonic restraints were applied to atoms in the buffer zone, scaled from a maximum at 25 Å to zero at a distance of 21 Å from the center. Force constants for different atom types were calculated from approximate average X-ray temperature factors. The outer "reservoir" zone contained all atoms further than 25 Å from the center. No cutoff was applied for the nonbonded interactions in the QM/MM calculations.

To test the QM/MM model with respect to its ability to maintain the correct structure of QM region in the RT active site, two different semiempirical methods (AM1 and PM3) (Dewar *et al.*, 1985; Dewar *et al.*, 1989; Stewart. 1989; Stewart. 1991; Hutter *et al.*, 1998) were applied for the QM region while the remainder of the system was treated by the MM (CHARMM27) (MacKerell *et al.*, 1998; MacKerell *et al.*, 1995; Foloppe *et al.*, 2000; Mackerell *et al.*, 2000) force field. QM/MM energy minimization (500 steps of SD and then 1750 steps of ABNR) was performed on minimized MM snapshot taken from the final MD structure (at 500 ps of MM MD simulation) in both deprotonated (Model A) and mono-protonated (Model B) forms of the triphosphate moiety. Then, the modelled systems were equilibrated for 1 ps, afterwards next 9 ps of molecular dynamics were performed in the small QM subsystem partition scheme. Where as the longer molecular dynamics (49 ps) were treated in the large QM subsystem partition scheme.
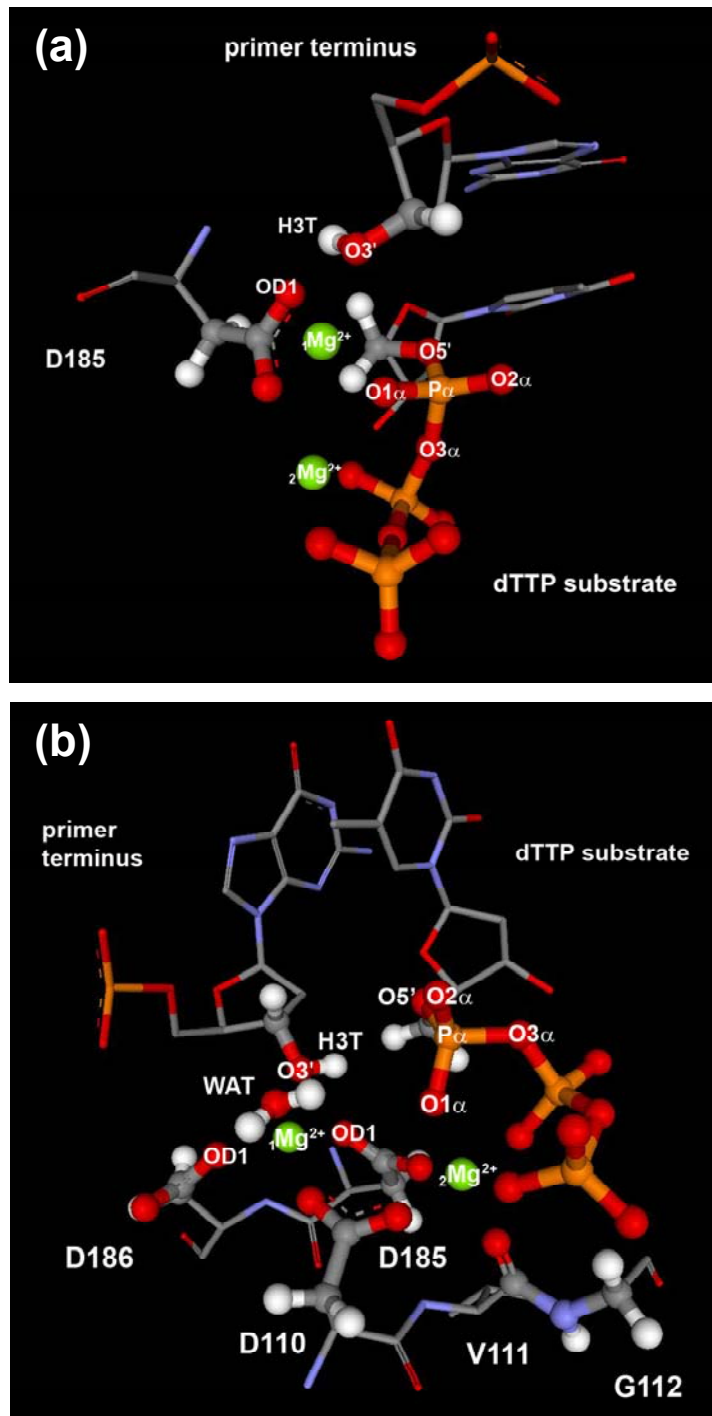
Figure 9 Two modelled structure represented for the active site: (a) the small QM/MM partition scheme; (b) the large QM/MM partition scheme. QM region is shown in style of bond and stick.

Two reaction coordinates of interest are the proton transfer (deprotonation) from the 3'-hydroxyl group of the terminal primer to an oxygen of H-acceptor and the nucleophilic attack from the O3'-hydroxyl oxygen of the terminal primer on the α-phosphorus of the dTTP substrate. The reaction coordinates for deprotonation were defined as the forming H3T-O bond, and the difference in length between the breaking O3'-H3T bond in the primer terminus and forming H3T-O bond (d(O3'-H3T) – d(H3T-O)) where an oxygen (O) of H-acceptor was proposed to be the OD1-carboxylate oxygen of Asp185, the OD1-carboxylate oxygen of Asp186 or the O5' oxygen of the α-phosphate. In nucleotide addition, the reaction coordinate were represented by the forming substrate Pα-O3' primer bond, and the difference in length between the breaking Pα-O3α bond in substrate and forming bond.

The next stage was harmonically restrained dynamics to a range of values along the reaction coordinate using umbrella sampling which the reaction coordinate was sampled every 0.1 or 0.2 Å. The force constants for umbrella potential were 200 and 300 kcal/mol in deprotonation and nucleotide addition, respectively, which resulted in simulations sampling sufficiently overlapping regions of the reaction coordinate. Reaction coordinate values were collected every time step. Each simulation is comprised of 1 ps of equilibration and then 9 ps of sampling dynamics in the small QM subsystem partion scheme. In contrast, the longer molecular dynamics (29 ps) were treated in the large QM subsystem partion scheme. Each subsequent equilibration is started from the 1-ps point of the adjacent equilibration run. Finally, all trajectory files with reaction coordinate values were analyzed using the weighted histogram analysis method (WHAM) (Kumar *et al.*, 1992; Bartels *et al.*, 1997), resulting in a free energy profile.
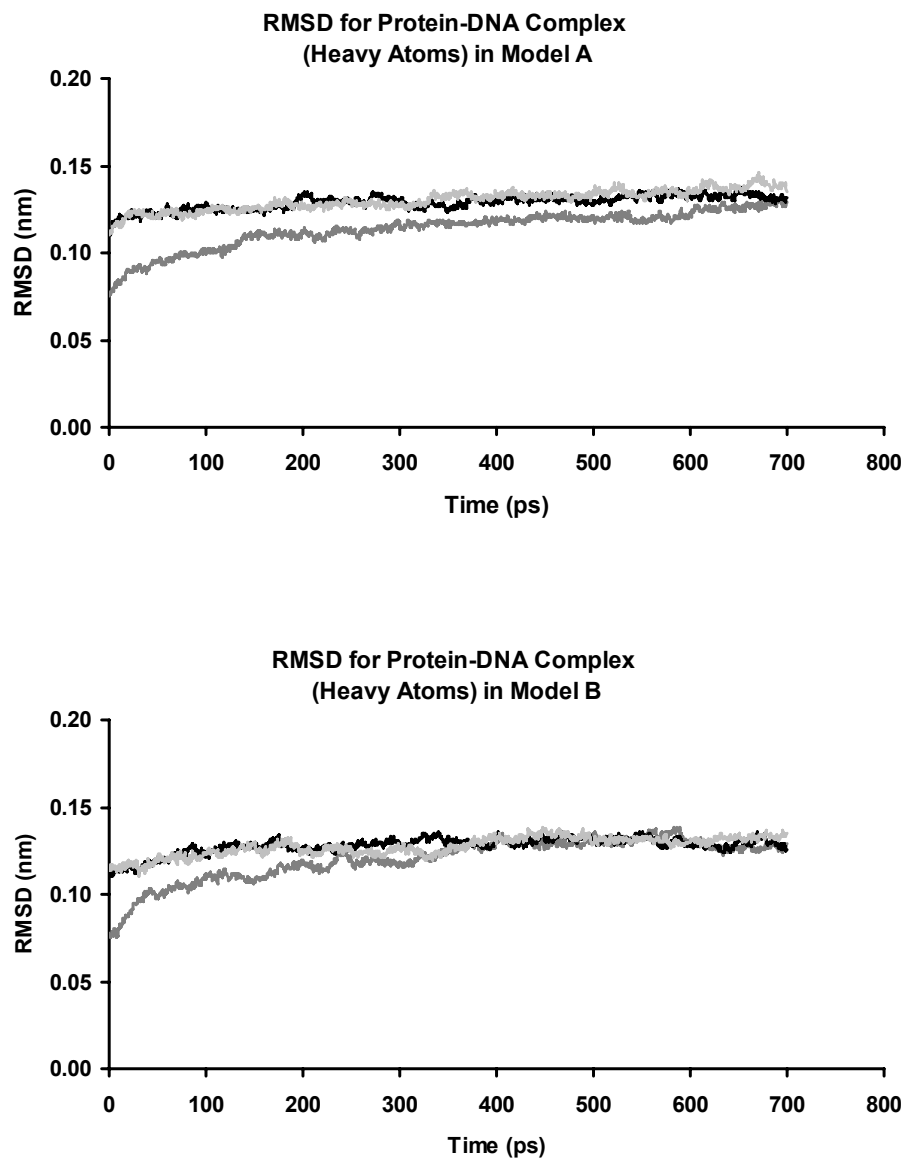
RESULTS AND DISCUSSION

Molecular Dynamics Investigation on the RT/dsDNA/dTTP ternary complex

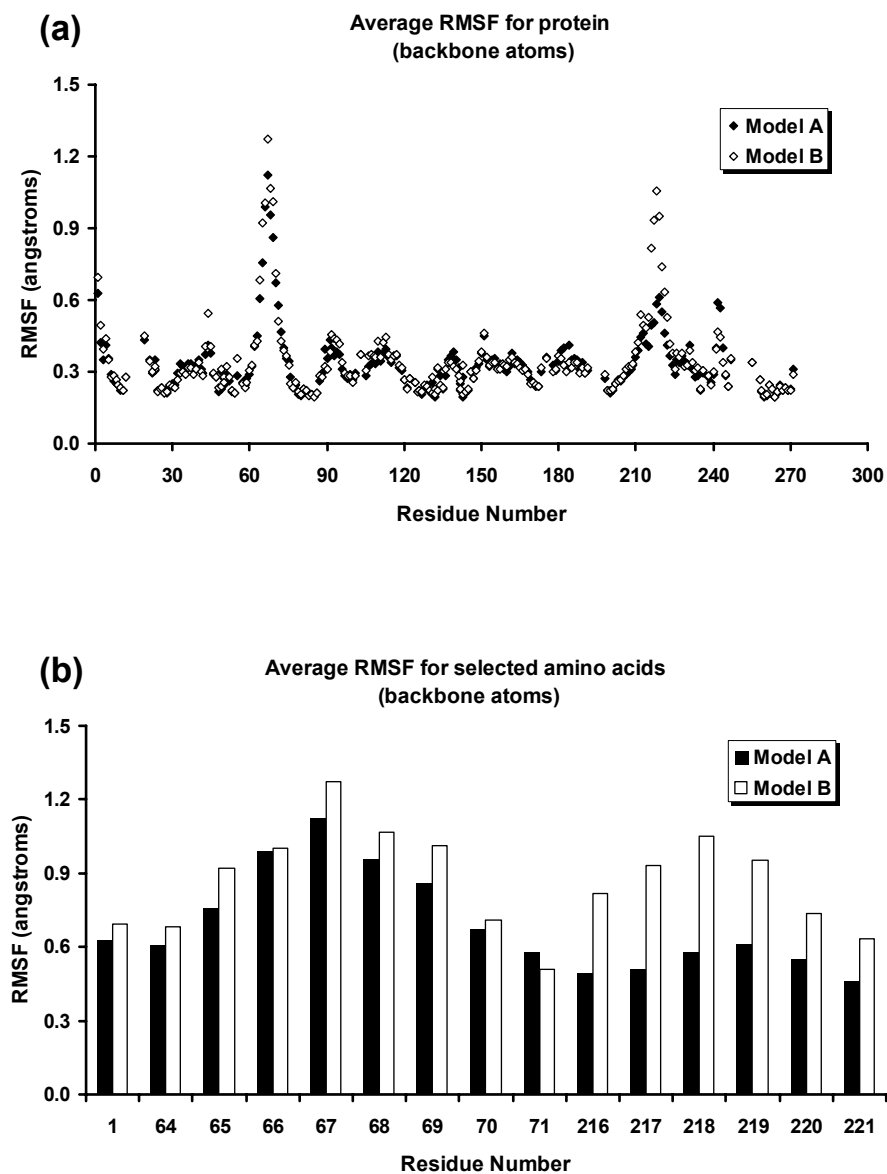**1. Molecular dynamics simulations**

To assess the stability of the simulations, the root-mean-square deviations (RMSD) between the structures obtained during the trajectories and the initial structure for all heavy atoms in the protein-DNA complex at picosecond intervals in the simulations are shown in Figure 10. As can be seen from the plot of RMSD versus simulation time for the two systems (Models A and B), the RMSD is steady after approximately the first 200 ps of the simulation, indicating a reasonably stable system. The RMSDs averaged over the last 3 500ps simulations for both models A and B are close to 0.13 nm which may not reveal any significant differences in the overall behavior that can be attributed to the two different protonation states of triphosphate moiety.

The root-mean-square fluctuation (RMSF) of protein backbone atoms for individual residues averaged over the 3 500ps simulations for two systems studied (Models A and B) is shown in Figure 11(a). The results indicate that the most flexible regions correspond to amino acids 64-71 (see Figure 11(b)) whose locations are close to polymerase active site. The presence of a hydrogen atom on the O2γ-oxygen of triphosphate moiety (Model B) particularly increases the flexibility of the residues 216-221 surrounding the dNTP binding pocket (Figure 11(b)) (e.g. the average RMSD of Asp218 is c.a. 0.58 Å: Model A and c.a. 1.05 Å: Model B). The fully deprotonated dTTP substrate model altogether shows less movement, suggesting a more stable active site region.

**RMSD for Protein-DNA Complex
(Heavy Atoms) in Model A**

**RMSD for Protein-DNA Complex
(Heavy Atoms) in Model B**

Figure 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.
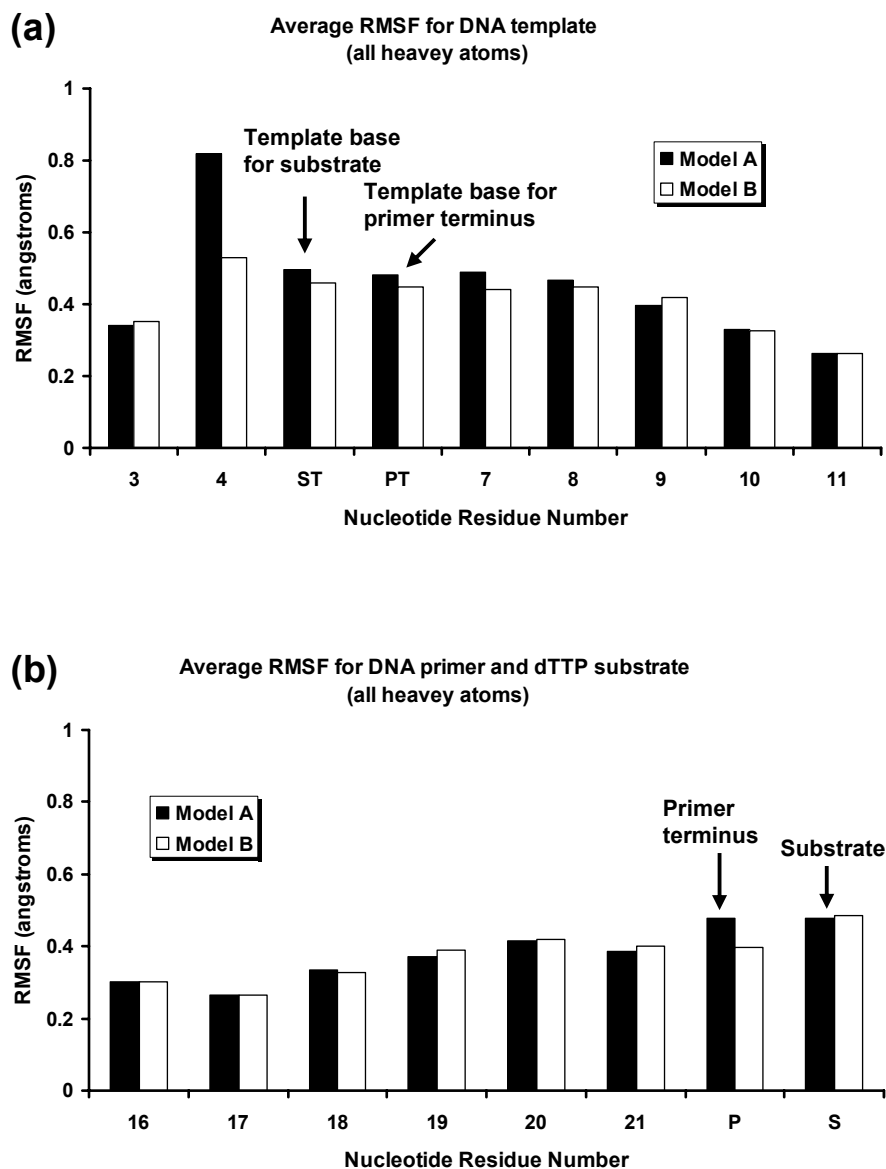
Source: Rungrotmongkol (2006)

(a)

**Average RMSF for protein
(backbone atoms)**

(b)

**Average RMSF for selected amino acids
(backbone atoms)**

Figure 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 3(a) and for some selected residues in 3(b).

Source: Rungrotmongkol (2006)

Figure 12 presents the average RMSFs of all heavy atoms for the double-stranded DNA fragment. Whereas most of nucleotides show comparable movement to each other in models A and B, the nucleotide at the position 4 (Cyt4) towards the 5' end of the DNA template, which codes for the next base of DNA primer, exhibits the most movement. This is particularly noticeable in the deprotonated triphosphate system (Model A). The nucleotides located in polymerase active site are the template base for substrate (ST) and primer terminus (PT) (Ade5 and Cyt6), and the primer terminus (P) (Gua22). These nucleotides (ST, PT and P) and the dTTP substrate (S) have similar movement; these suggested that these two simulated structures (Models A and B) both maintain stable active sites.

Table 2 compares the average RMSD of amino acids D110-Y115 and M184-D186, the part of the polymerase active site where the chemical reactions take place. The RMSD is calculated between the average structure of saved structures from all trajectories for a single simulation and the crystal structure of the HIV-1 RT/DNA/dTTP ternary complex. The average RMSD is evaluated between the three RMSDs in each model. It is noted that the RMSDs are similar for all different simulations as shown in small standard deviations (< 0.1 Å). The average RMSDs for the protein backbone atoms and for all protein heavy atoms are relatively small comparable in both simulated models. It is seen that these amino acids at the active site have similar conformations in the two structures simulated and in the crystal structure. The active site structure is maintained well in the simulations.

(a)

Average RMSF for DNA template
(all heavey atoms)

(b)

Average RMSF for DNA primer and dTTP substrate
(all heavey atoms)

Figure 12 Root mean square fluctuations (RMSF) (Å) of all heavy atoms by nucleotide averaged over 3 500 ps MD simulations for the DNA template [4(a)], DNA primer and dTTP substrate (S) [4(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.

Source: Rungrotmongkol (2006)

Table 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 Oγ2-oxygen) with respect to the crystal structure of the HIV-1 RT/DNA/dTTP ternary complex[a] (Huang *et al.*, 1998)

Model	Average RMSD (Å)	
	Protein Backbone Atoms	All Protein Heavy Atoms
Model A	0.44 (0.05)	0.78 (0.06)
Model B	0.62 (0.08)	0.95 (0.05)

[a] Standard deviations (δ) in Å is reported in parentheses.

Table 3 Selected structural parameters in Models A and B averaged over 3 500 ps MD simulations for each model[a,b]

	Model A	Model B
Mg1 coordination sphere:		
D[Mg1-O3'(primer)]	2.1(0.1)	2.1(0.1)
D[Mg1-O1α(dTTP)]	3.1(0.2)[0.1]	3.1(0.2)[0.1]
D[Mg1-OD2(D110)]	1.8(0)[-0.9]	1.8(0)[-0.9]
D[Mg1-OD1(D185)]	1.8(0)[-0.2]	1.8(0)[-0.2]
D[Mg1-OD1(D186)]	1.8(0)[-1.9]	1.8(0)[-1.9]
D[Mg1-WAT]	1.9(0)	1.9(0)
Mg2 coordination sphere:		
D[Mg2-O1α]	1.9(0.1)[-0.3]	1.9(0.1)[-0.3]
D[Mg2-O2β]	1.9(0.1)[-0.3]	1.9(0.1)[-0.3]
D[Mg2-O1γ]	1.8(0)[-0.6]	1.9(0)[-0.5]
D[Mg2-OD1(D110)]	1.9(0)[-0.2]	1.8(0)[-0.3]
D[Mg2-O(V111)]	2.1(0.1)[-0.2]	2.0(0.1)[-0.3]
D[Mg2-OD2(D185)]	1.9(0)[-0.5]	1.8(0)[-0.6]
Other active site distances:		
D[Mg1-Mg2]	4.0(0.1)[0.4]	3.9(0.1)[0.3]
D[Pα(dTTP)-C3'(primer)]	4.6(0.2)[-0.1]	4.4(0.1)[-0.3]
D[Pα(dTTP)-O3'(primer)	3.6(0.1)	3.5(0.1)
D[OD2(D185)-O1α(dTTP)]	2.6(0.1)[-0.1]	2.7(0.1)[0]
D[OD2(D185)-O2β(dTTP)]	2.8(0.1)[0]	2.7(0.1)[-0.1]
D[OD2(D186)-O1P(primer)]	4.1(0.3)[0.3]	3.9(0.2)[0.1]
D[H3T(primer)-O5'(dTTP)]	**1.9(0.1)**	**1.9(0.1)**
D[HZ1(K65)-O3γ(dTTP)]	2.9(0.8)	2.1(0.5)

Table 3 (cont'd)

	Model A	Model B
D[O3γ(dTTP)-NZ(K65)]	3.2(0.5)[0.5]	2.8(0.1)[0.1]
D[HZ3(K65)-O3β(dTTP)]	2.9(0.9)	3.3(0.4)
D[O3β(dTTP)-NZ(K65)]	3.2(0.5)[-0.2]	3.8(0.3)[0.4]
D[NH2(R72)-NZ(K65)]	**4.4(0.4)[-0.1]**	**5.5(0.6)[1.0]**
D[HH21(R72)-O2α(dTTP)]	**3.0(0.5)**	**5.0(1.0)**
D[HH22(R72)-O2α(dTTP)]	**4.4(0.5)**	**4.2(0.4)**
D[NH2(R72)- O2α(dTTP)]	**3.9(0.4)[0.9]**	**4.9(0.6)[1.9]**
D[HH22(R72)-O1β(dTTP)]	**3.1(0.3)**	**6.1(0.5)**
D[NH2(R72)-O1β(dTTP)]	**3.6(0.4)[-0.8]**	**7.0(0.7)[2.6]**
D[HH21(R72)-O3α(dTTP)]	**2.6(0.3)**	**4.6(0.6)**
D[NH2(R72)-O3α(dTTP)]	**3.2(0.2)[0.5]**	**5.5(0.7)[2.8]**
D[HN(D113)-O3β(dTTP)]	3.4(0.2)	2.2(0.3)
D[N(D113)-O3β(dTTP)]	4.3(0.2)[1.3]	3.2(0.3)[0]
D[HN(A114)-O2β(dTTP)]	2.3(0.2)	2.5(0.2)
D[N(A114)-O2β(dTTP)]	3.3(0.2)[0.1]	3.3(0.2)[0.1]
D[H3T(dTTP)-O1β(dTTP)]	**2.4(0.9)**	**1.7(0.1)**
D[O3'(dTTP)-O1β(dTTP)]	**3.2(0.5)[0.2]**	**2.6(0.2)[-0.4]**
D[H3T(dTTP)-O2β(dTTP)]	3.7(0.8)	3.5(0.2)
D[O3'(dTTP)-O2β(dTTP)]	4.2(0.4)[0.1]	4.2(0.2)[0.1]
D[C3'(dTTP)-Pγ(dTTP)]	6.6(0.1)[-0.1]	6.4(0.1)[-0.3]
D[H3T(primer)-O5'(dTTP)]	1.8(0.1)	1.9(0.1)
D[O3γ(dTTP)-H(wat5347)]	**2.4(0.7)**	**3.5(1.9)**
D[O3γ(dTTP)-H(wat7945)]	**2.4(0.6)**	**3.6(1.9)**
Angles:		
A[O3'(primer)-H3T--O5'(dTTP)]	160(10)	161(10)
A[NZ(K65)--HZ1--O3γ(dTTP)]	87(37)	136(36)
A[NZ(K65)--HZ3--O3β(dTTP)]	106(44)	116(21)
A[C3'--Pα--Pβ(dTTP)]	74(4)[12]	68(2)[6]

[a] Averaged distances (d), deviations (δ), and angles (∠) from the starting and X-ray structures are reported for Models A and B. Distances and deviations are in Å, angles in degree. Standard deviation and deviation from the X-ray structure are reported in parentheses and square brackets, respectively.

[b] Atoms in the active site of the models are labeled as the corresponding atoms in crystal structure in the case of added 3'OH-primer and TIP3P water molecules.
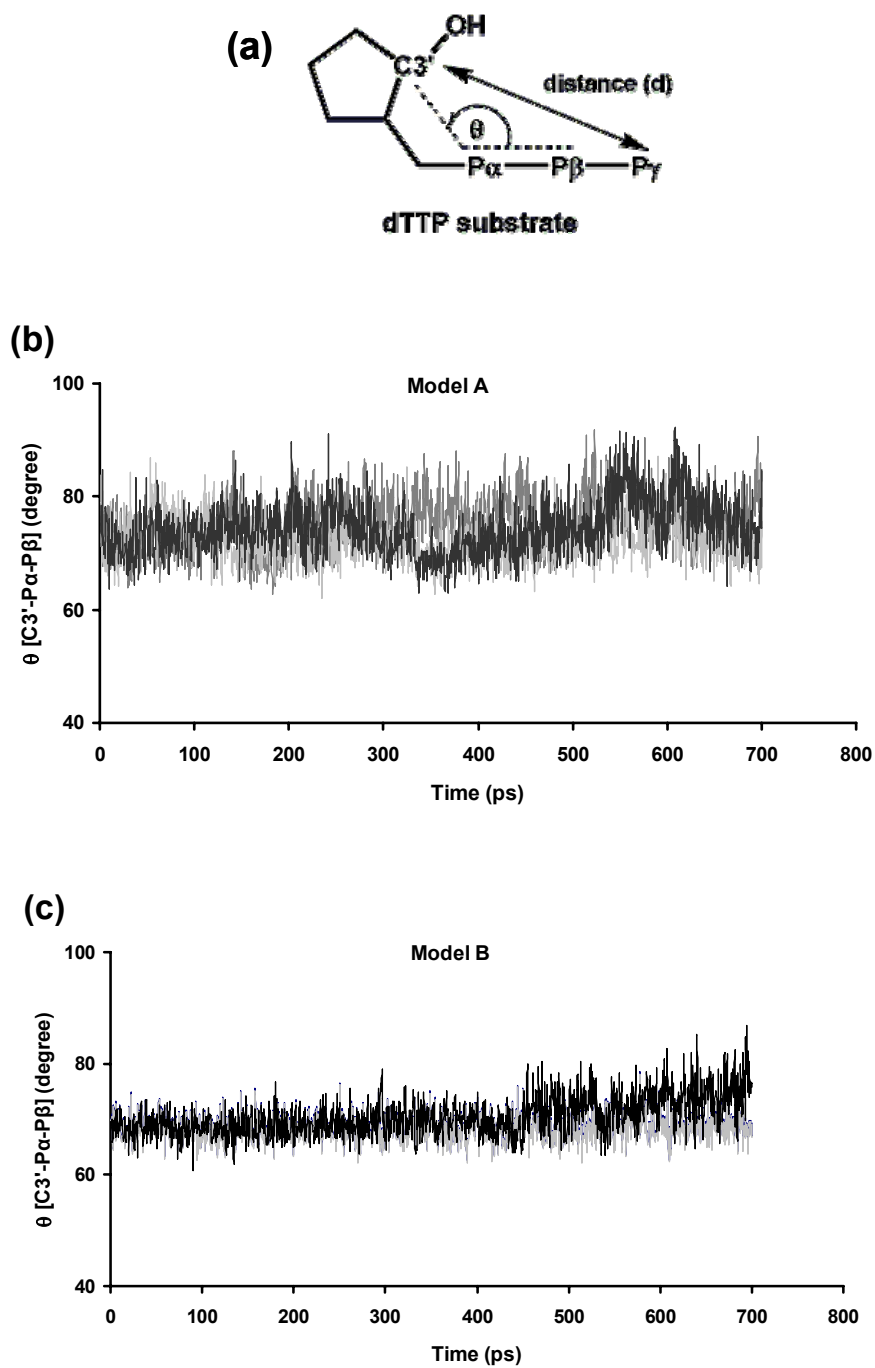
Figure 13   (a) Definition of angle θ and distance d. Plot of angle θ as a function of time in Models A (b) and B (c). Plot of distance as a function of time in Models A (d) and B (e). Three 700 ps MD simulations are shown in the plot with different shades.

(d)

(e)
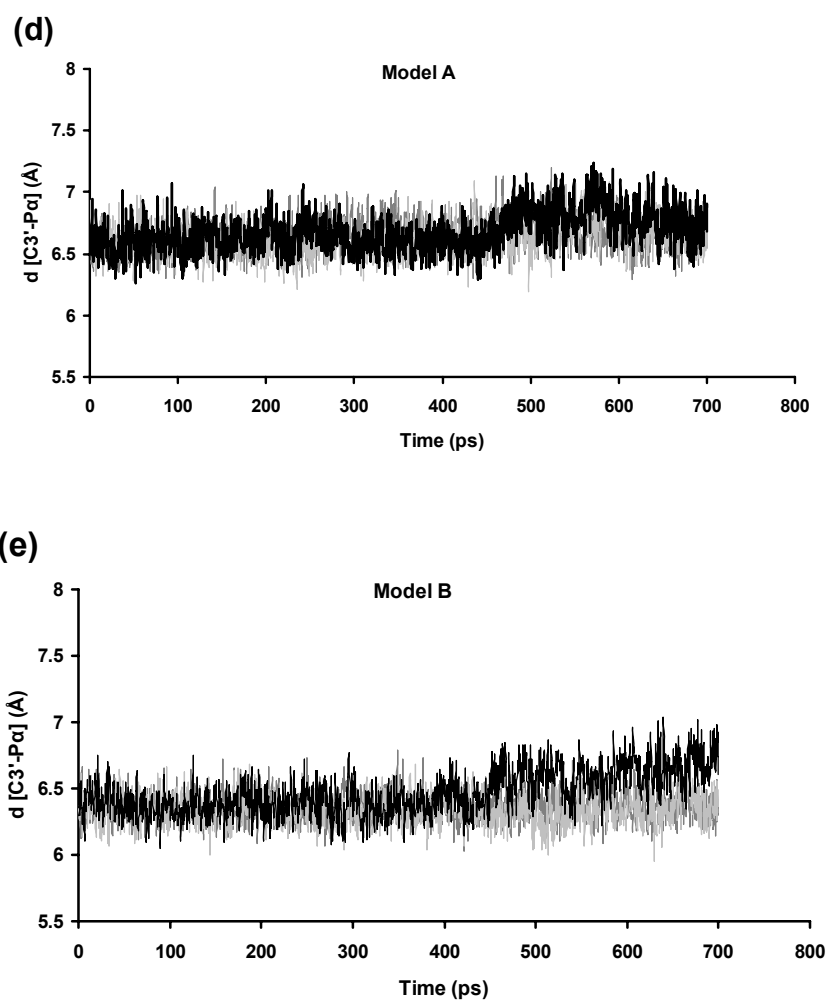
Figure 13   (cont'd)

2. Coordination of the two magnesium ions, and the dNTP binding site

The average values of some important structural parameters (averaged over the 3 500ps simulations) in models A and B are shown in Table 3. The simulations of both the deprotonated (Model A) and the mono-protonated (Model B) forms of the triphosphate moiety showed similar coordination of the catalytic and the nucleotide-bridging magnesium ions. The largest structural changes in the catalytic ion ($_1Mg^{2+}$) coordination were due to the presence of the 3'-OH terminal primer which both simulated systems provided the orientation of the 3'-OH group directly toward the $_1Mg^{2+}$ ion (the average distance D[Mg1-O3'(primer)] is 2.1 Å). Other changes in the coordination of $_1Mg^{2+}$ include an added water molecule (D[Mg1-WAT] = 1.9 Å), and the acetate sidechain of D186 rotated to be a new ligand (D[Mg1-OD1(D186)] = 1.8 Å). It should be noted that the D186 carboxylate group is observed to coordinate much more tightly to the catalytic magnesium ion in the model A and B simulations than in the crystal structure (3.7 Å). In the coordination environments observed in the simulations (Models A and B), two bridging carboxylate groups (of the D110 and D185 aspartates) are positioned symmetrically between two magnesium ions. In contrast, only bridging D185 carboxylate group is coordinated symmetrically to the divalent cations whereas the D110 carboxylate group is twisted toward the $_1Mg^{2+}$ ion, with both carboxylate oxygen atoms coordinated to this magnesium ion, and the OD1 carboxylate oxygen acts as a bridging ligand of two magnesium ions found in the crystal structure. The $_1Mg^{2+}$ ligands altogether therefore are the sidechain of three aspartates (D110, D185 and D186), the bridging O1α oxygen of the α-phosphate, the 3'-OH terminal primer and a modelled water molecule arranged approximately octahedral which these interactions are not observed in the crystal structure [36]. The octahedral character of the $_2Mg^{2+}$ coordination sphere observed in the crystal structure is retained. Its original six oxygen ligands, the O1α bridging oxygen and two nonbridging oxygen atoms (O2β and O1γ) on the triphosphate moiety, the carbonyl oxygen of V111, and the OD1 oxygens of D110 and D185, move a little closer to the $_2Mg^{2+}$ ion in the model A and B simulations. Considering on the reaction coordinate of nucleotide addition, the average distance between the α-phosphorus of the dTTP substrate and the 3'-hydroxyl carbon on the DNA primer terminus was slightly shorter

(by 0.1 Å: Model A and 0.3 Å: Model B) than that obtained in the crystal structure (4.7 Å), accordingly the average Pα-O3' distance was closed to 3.6 Å: Model A and 3.5 Å: Model B.

Figure 13 shows the plots of angle θ between atoms C3'-deoxyribose carbon, Pα- and Pβ-triphosphate phosphorus, and of distance d between C3'-deoxyribose and Pγ-triphosphate phosphorus atoms on dTTP. With the small fluctuations of the deoxyribose ring with respect to the triphosphate moiety observed from angle θ and distance d in both the deprotonated and mono-protonated forms of dTTP triphosphate systems, thus, it presents the high triphosphate conformational stability during MD simulations probable due to the stabilizations from some amino acid environment and bound water molecules which will be discussed more details in the next section. Obviously, the average angle θ and distance d are given in Table 3; these are 74 degree and 6.6 Å in Model A; and 68 degree and 6.4 Å in Model B.

In contrast, Model C, mono-protonated on the O3γ atom, showed a distorted triphosphate geometry of dTTP (Figure 14). It can be seen that γ-phosphate moiety moved its position to be closer to the β-phosphate leading to the apparent formation of a P_β-P_γ (2.4 Å) bond. Consequently, both magnesium ions move away from the terminal DNA primer and dTTP, losing their O1α bridging oxygen and 3'-OH primer ligands. Model B appears the more appropriate mono-protonated triphosphate form of the dTTP substrate. The results with Model C did not appear realistic. Hence, only Models A and B will be further discussed.
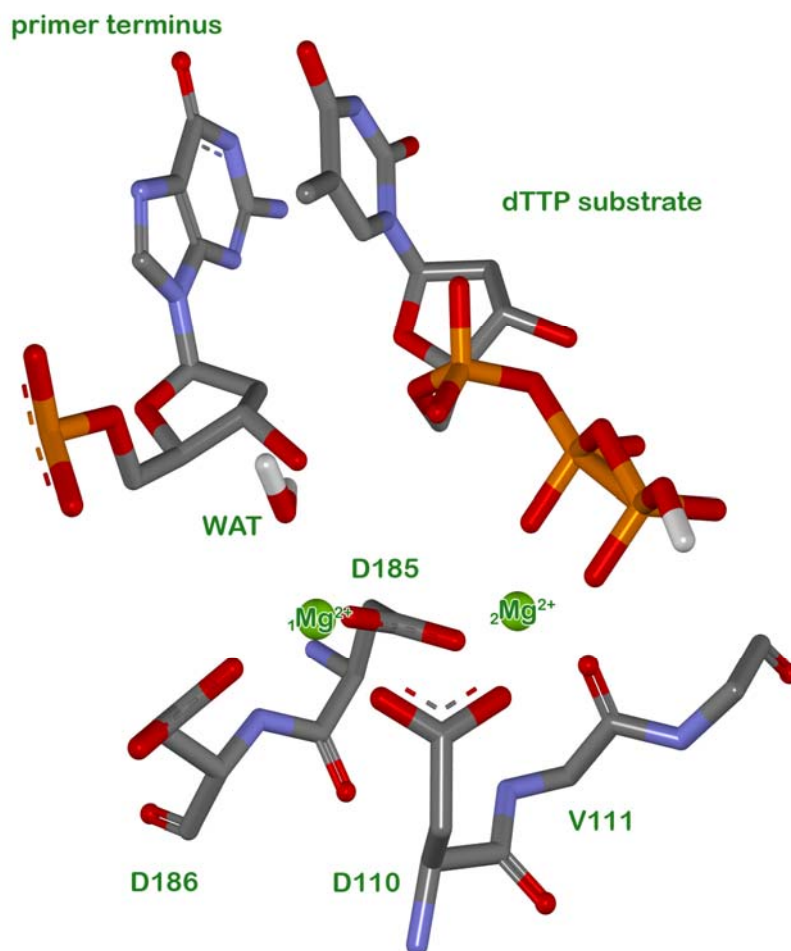
Figure 14   Conformation of mono-protonated triphosphate on Oγ3 oxygen of dTTP substrate, two Mg^{2+} ions and their coordination environments (Model C) in the final structure from MM MD simulation. Hydrogen atoms are not shown except for the mono-protonated O3γ atom and the bound water molecules. The colours were used by atom type colour.

Source: Rungrotmongkol 2006

Table 4 Hydrogen bond frequencies for protein– and water–phosphate hydrogen bonds averaged over 3 500 ps MD simulations in Models A and B[a]

Hbond pairs	HbF	
	Model A	Model B
O2α-HH21(R72)	0.36 (0.35)	-
O2α-wat4006	0.04 (0.03)	0.60 (0.60)
O2α-wat5347	0.70 (0.70)	0.68 (0.65)
O2α-wat7572	0.36 (0.36)	-
O2α-wat7724	0.90 (0.90)	-
O3α-HH21(R72)	0.22 (0.10)	0.01 (0.01)
O3α-HH22(R72)	0.26 (0.22)	-
O3α-H3T(dTTP)	0.38 (0.29)	-
O3α-wat3221	-	0.38 (0.36)
O1β-HH21(R72)	0.10 (0.03)	-
O1β-HH22(R72)	0.03 (0)	-
O1β-HN(A114)	0.18 (0.18)	0.68 (0.68)
O1β-H3T(dTTP)	0.92 (0.92)	0.99 (0.99)
O1β-wat7281	0.90 (0.90)	-
O1β-wat7623	-	0.86 (0.86)
O2β-HN(D113)	-	0.02 (0.02)
O2β-HN(A114)	0.84 (0.84)	0.60 (0.59)
O3β-HZ1(K65)	0.24 (0.22)	0.02 (0.02)
O3β-HZ2(K65)	0.11 (0.11)	0.01 (0.01)
O3β-HZ3(K65)	0.43 (0.41)	-
O3β-HH22(R72)	0.02 (0.02)	-
O3β-HN(D113)	0.01 (0)	0.87 (0.87)
O3β-wat5347	0.59 (0.34)	-
O1γ-wat6565	0.85 (0.81)	-
O1γ-wat7945	0.33 (0.19)	-
O2γ-HZ1(K65)	-	0.06 (0.06)
O2γ-HZ3(K65)	0.01 (0.01)	-
O2γ-HN(D113)	0.99 (0.99)	0.20 (0.20)
O2γ-wat2947	0.42 (0.42)	-
O2γ-wat7350	0.72 (0.72)	-
O2γ-wat7576	0.38 (0.37)	-

Table 4 (cont'd)

Hbond pairs	HbF	
	Model A	**Model B**
O3γ-HZ1(K65)	0.20 (0.16)	0.84 (0.83)
O3γ-HZ2(K65)	0.13 (0.06)	0.14 (0.14)
O3γ-HZ3(K65)	0.50 (0.38)	-
O3γ-wat5347	0.74 (0.71)	0.51 (0.51)
O3γ-wat7011	0.32 (0.32)	-
O3γ-wat7623	0.39 (0.39)	-
O3γ-wat7945	0.56 (0.55)	-

[a] Hydrogen-bond frequency (HbF) was calculated as the number of hydrogen bonds for each set of interacting atoms divided by the total number of saved structures from all trajectories for a particular model. HbF = 1 means a hydrogen bond constantly maintained during the 3 500 ps MD simulations. Our selection of hydrogen bonds was based on two criteria: [I]. d(hydrogen--H-bond acceptor) ≤ 2.5 Å and H-bond donor--hydrogen--H-bond acceptor) $\geq 90°$, and [II]. d(hydrogen--H-bond acceptor) ≤ 2.5 Å and (H-bond donor--hydrogen--H-bond acceptor) $\geq 120°$ as shown in parentheses [64].

Hydrogen bonds in the substrate itself, and between the substrate and the enzyme environment or with bound water molecules, play important roles in substrate binding at the active site. In Table 3, the intramolecular hydrogen bonding in the substrate between its deoxyribose ring and O1β oxygen (the average distances D[H3T(dTTP)-O1β(dTTP)] and D[O3'(dTTP)-O1β(dTTP)]) are estimated in the model A and B simulations. It was found that the hydrogen bond is stronger in Model B (where the average hydrogen bond distances are 1.7 Å and 2.6 Å, respectively), than in Model A (where they are 2.4 Å and 3.2 Å, respectively). The deprotonated triphosphate moiety in Model A was found to have some hydrogen-bonding interactions with the methylammonium of K65, the guanidinium of R72 and the backbones (NHs) of D113 and A114. In contrast, the hydrogen bonds with R72 are not preserved in Model B: the guanidinium sidechain of R72 moves away from the mono-protonated triphosphate. During the three 500 ps molecular dynamics trajectories of Model A and B, the dTTP substrate is stabilized not only by the amino

acids in the dNTP binding pocket, but also by some tightly bound water molecules (modelled waters) interacting extensively with its triphosphate moiety, as indicated by bold values shown in Table 3.

To analyze these hydrogen bonds, the hydrogen-bond frequency (HbF) was calculated as the number of hydrogen bonds (where the definition of hydrogen bond is given in Table 4) for each set of interacting atoms divided by the total number of saved structures from all trajectories for a particular model. HbF = 1 means a hydrogen bond constantly maintained during the 3 500 ps MD simulations. The results are given in Table 4. These results show for what proportion of a simulation a particular hydrogen bond is present. HbFs for protein-phosphate in two systems studied (Models A and B) are compared in Table 4: these are the hydrogen bonds between triphosphate oxygens and active site residues (the sidechain of K65 and R72, and the backbone of D113 and A114). It can be seen that majority of HbF values (e.g. HbF[O2α-HH21(R72)] is 0.36: Model A and 0: Model B; HbF[O2β-HN(A114)] is 0.84: Model A and 0.60: Model B) are higher in Model A. Also most of the oxygens on the triphosphate groups have hydrogen bonds with bound water molecules, with the HbF values larger than 0.30, mostly found in the simulations of deprotonated dTTP. These indicated that phosphate appears more tightly bound in Model A which might be expected given its higher negatively charge in this model. The results altogether supported the model of the deprotonated form of dTTP (Model A) to be more stable than the model of the mono-protonated form of dTTP (Model B) and its orientation was in good agreement with the experimental structure (Huang *et al.*, 1998). The nature of the intermolecular interactions between HIV-1 RT and the nucleosidic phosphate has also been studied by *ab initio* molecular dynamics calculations for 1.5 ps (Alber *et al.*, 2000) which gave similar results; these were the intra- and inter-molecular hydrogen bonding interactions and the hydrogen-bond frequencies (with the Arg72 and Lys65 sidechains, the Asp113 and Ala114 backbones and the modelled waters) mostly found in the deprotonated triphosphate of dTTP.

3. Protonation states of the aspartic sidechain

We considered and tested several possibilities for protonation of two aspartic acids, D185 and D186, by adding a proton on the appropriate carboxylate oxygen and then performing molecular dynamics simulations for 700 ps to determine the effects of the protonation.

Table 5 showed the active-site atom distances averaged over the last 500 ps MD trajectories of four different systems. Simulation of the protonated D185 systems (HD185-Model A and HD185-Model B), produced similar results to one another for the structural reactive part of an attacking nucleophile and the magnesium coordination sphere. The average distance between the reacting α-phosphorus of the dTTP substrate and the 3'-hydroxyl oxygen of the primer terminus (3.3 Å: HD185-Model A and 3.2 Å: HD185-Model B) was slightly shorter than in the Model A and B simulations (3.6 Å: Model A and 3.5 Å: Model B) (see Table 5 and 2). The Mg1-O1α distance decreased significantly to 1.9 Å (compared to 3.0 Å in the 1RTD crystal structure and to 3.1 Å in the previous Model A and B results) while the average Mg1-O3' distance was equal to 2.2 Å, with the catalytic magnesium ion tightly close to the α-phosphate groups and the 3'-OH terminal primer. The distance between two magnesium ions increased significantly to 4.7 Å (the standard deviation = 0.1 Å) (compared to 3.6 Å in the crystal structure). These simulations also showed reorganization in the first coordination shell of the nucleotide-binding magnesium ion: the $_2Mg^{2+}$ ion lost two of its ligands, the O1α-phosphate oxygen of the dTTP triphosphate (the average distance D[Mg2-O1α] is 3.9 Å) and the OD2-carboxylate oxygen of D185 (D[Mg2-OD2(D185)] = 4.1 Å), compensating with other two modelled water molecules (each average distance D[Mg2-WAT] is 2.0 Å). Both protonated D185 systems have hydrogen bonds in place between the OD2-carboxylate oxygen of D185 and two phosphate oxygens (O1α and O2β) of the dTTP substrate, as indicated by bold values in Table 5.

Table 5 Atom distances (Å) 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[a,b]

	HD185-Model A	HD185-Model B	HD186-Model A	HD186-Model B
Mg1 coordination sphere:				
D[Mg1-O3'(primer)]	2.2(0.1)	2.2(0.1)	3.2(1.0)	4.4(0.2)
D[Mg1-O1α(dTTP)]	1.9(0.1)[-1.1]	1.9(0.1)[-1.1]	3.7(0.6)[0.7]	1.8(0.0)[-1.2]
D[Mg1-OD2(D110)]	1.8(0)[-0.9]	1.8(0)[-0.9]	1.8(0)[-0.9]	1.8(0)[-0.9]
D[Mg1-OD1(D185)]	2.2(0.1)[0.2]	2.2(0.2)[0.2]	1.8(0)[-0.2]	1.8(0)[-0.2]
D[Mg1-OD1(D186)]	3.4(0.2)[-0.3]	1.8(0)[-1.9]	3.6(1.6)[-0.1]	2.1(0.1)[-3.1]
D[Mg1-OD2(D186)]	1.8(0)[-3.4]	3.3(0.2)[2.0]	4.7(0.8)[0.5]	4.1(0.2)[0.4]
D[Mg1-WAT4006]	1.9(0.1)	1.9(0.1)	1.9(0.1)	2.0(0.1)
D[Mg1-WAT5884]			2.0(0.1)	2.0(0.1)
Mg2 coordination sphere:				
D[Mg2-O1α]	3.9(0.1)[1.7]	3.9(0.1)[1.7]	1.9(0.1)[-0.3]	3.9(0.1)[1.7]
D[Mg2-O2β]	1.9(0)[-0.3]	1.9(0)[-0.3]	1.9(0.1)[-0.3]	1.8(0)[-0.4]
D[Mg2-O1γ]	1.8(0)[-0.6]	1.8(0)[-0.6]	1.8(0)[-0.6]	1.9(0)[-0.5]
D[Mg2-OD1(D110)]	1.8(0)[-0.3]	1.8(0)[-0.3]	1.9(0)[-0.2]	1.8(0)[-0.3]
D[Mg2-O(V111)]	2.1(0.1)[-0.2]	2.0(0.1)[-0.3]	2.0(0.1)[-0.3]	2.0(0.1)[-0.3]
D[Mg2-O(G112)]				2.2(0.1)[-3.0]
D[Mg2-OD2(D185)]	4.1(0.2)[1.7]	4.1(0.2)[1.7]	1.9(0)[-0.5]	1.8(0)[-0.6]
D[Mg2-WAT6548]	2.0(0)	2.0(0)		
D[Mg2-WAT7945]	2.0(0.1)	2.0(0)		
Other active site distances:				
D[Mg1-Mg2]	4.7(0.1)[1.1]	4.7(0.1)[1.1]	4.1(0.3)[0.5]	4.2(0.1)[0.6]
D[Pα(dTTP)-C3'(primer)]	4.0(0.1)[-0.7]	4.2(0.2)[-0.5]	4.7(0.2)[0]	5.9(0.2)[1.2]
D[Pα(dTTP)-O3'(primer)	3.3(0.1)	3.2(0.1)	3.7(0.2)	5.7(0.2)
D[OD2(D185)-O1α(dTTP)]	3.2(0.2)[0.5]	3.3(0.2)[0.6]	2.6(0.1)	3.9(0.2)
D[HD2(D185)-O1α(dTTP)]	**2.4(0.2)**	**2.6(0.2)**		
D[OD2(D185)-O2β(dTTP)]	2.9(0.1)[0.1]	2.9(0.2)[0.1]	2.8(0.1)[0]	2.6(0.1)[-0.2]
D[HD2(D185)-O2β(dTTP)]	**2.1(0.2)**	**2.1(0.2)**		
D[OD2(D186)-O1P(primer)]	4.8(0.3)[1.0]	4.0(0.3)[0.2]	3.2(0.7)[-0.6]	3.8(0.7)[0]
D[HD2(D186)-O1P(primer)]			2.6(1.0)	3.6(0.9)

[a] Averaged distances (d), deviations (δ), and angles (∠) from the starting and X-ray structures are reported for Models A and B. Distances and deviations are in Å, angles in degree. Standard deviation and deviation from the X-ray structure are reported in parentheses and square brackets, respectively.

[b] Atoms in the active site of the models are labeled as the corresponding atoms in crystal structure in the case of added 3'OH-primer and TIP3P water molecules.
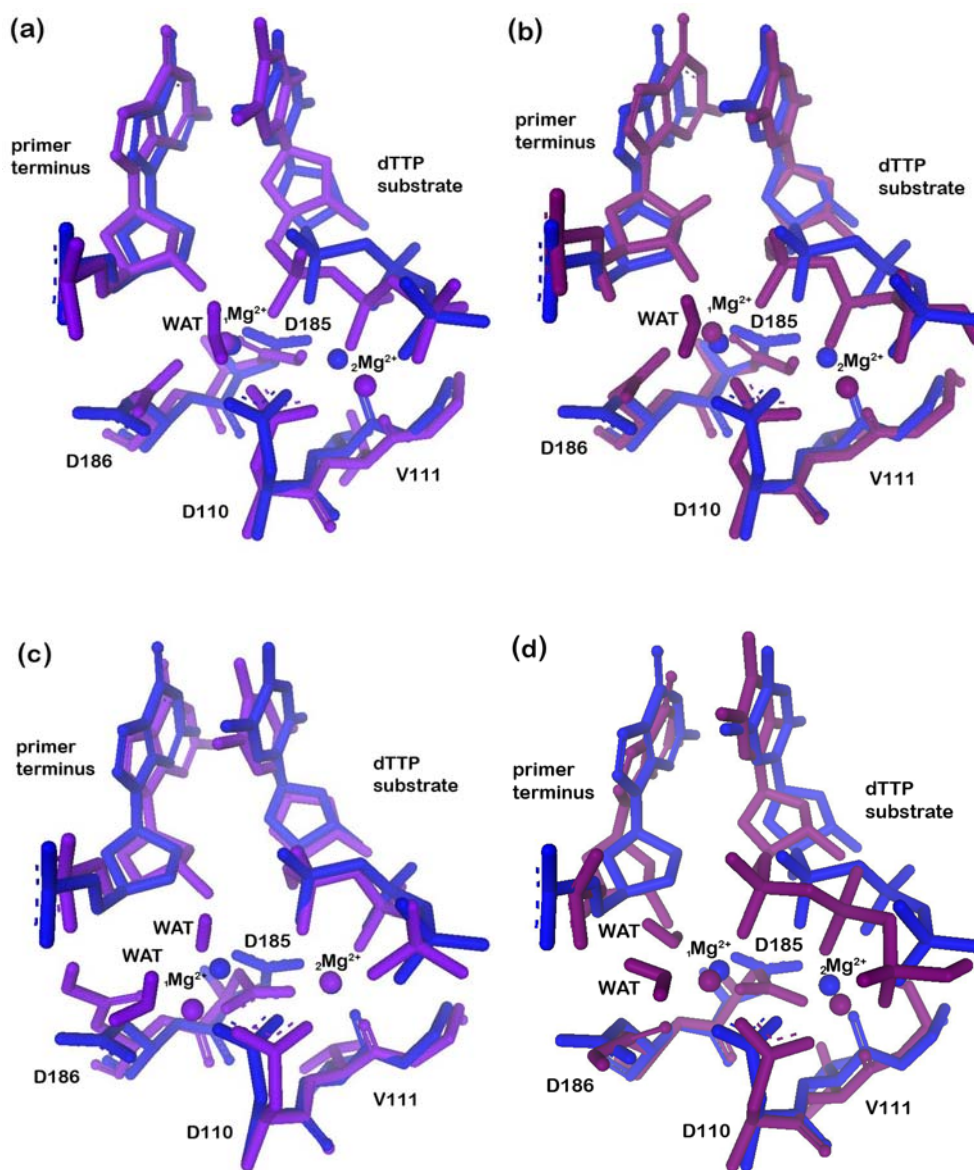
Figure 15   Superimposition of minimized CHARMM at last MD snapshot (Model A systems: blue-violet, Model B systems: red-violet) compared to the crystal structure (blue): 6(a) HD185-Model A; 6(b). HD185-Model B; 6(c) HD186-Model A; 6(d). HD186-Model B. Hydrogen atoms are not shown except for the mono-protonated O2γ atom and the bound water molecules.

Consequently, the average OD2(D185)-O1α distance increased (to 3.2 Å: HD185-Model A and 3.3 Å: HD185-Model B) from the crystal structure (2.7 Å) and the previous models A and B (2.6 Å: Model A and 2.7 Å: Model B) whereas the average OD2(D185)-O2β distance (2.9 Å: HD185-Models A and B) was constantly closed to the crystal (2.8 Å) and their original modelled structures (2.8 Å: Model A and 2.7 Å: Model B).

On the other hand, simulations of HD186-Model A (fully deprotonated dTTP triphosphate and the non-bridging D186 residue in the acidic form), showed the catalytic magnesium ion far from two coordinated oxygen ligands (the O1α and the 3'-hydroxyl oxygen atoms) (Mg1-O1α and Mg1-O3' distances closely to 3.7 Å and 3.2 Å given in Table 5). This finding indicated poorer agreement of this Model with the crystal structure. The rotation of the D186 carboxylate sidechain toward the $_1Mg^{2+}$ ion observed in Models A and B was not found in this Model. Therefore, coordinating the catalytic $_1Mg^{2+}$ as ligands are the sidechains of two aspartates (D110 and D185) and two modelled water molecule (coordination number = 4). A hydrogen bond interaction was found between the protonated aspartic acid (D186) and the phosphate linkage of primer terminus (the average distance D[HD2-O1P] = 2.6 Å) which the average OD2-O1P distance decreased (to 3.2 Å) from the crystal structure (3.8 Å) and the Model A results (4.1 Å) (Table 3). In other respects, HD186-Model A produced similar results to Models A and B. The simulation of HD186-Model B (mono-protonated triphosphate model with D186 protonated) showed clear changes in the ligands of the divalent cations. In this simulated structure, the six ligands of the $_1Mg^{2+}$ ion are the α-phosphate group (O1α), three active-site aspartic acid residues (D110, D185 and D186) and two modelled water molecules. The coordination of the other magnesium ion changed significantly from the experimental structure: the $_2Mg^{2+}$ ion was coordinated to the triphosphate moiety (only the O2β- and O1γ-oxygen atoms), the carbonyl oxygen of V111 and G112, and the acetate sidechain of D110 and D185. Not only did some of the magnesium ligands shift position, but also the average Pα-C3' distance was significantly longer (5.9 Å) than that obtained in the crystal structure (4.7 Å), consequently the average Pα-O3' distance (involved in nucleotide addition) increased (to 5.7 Å) relative to the previous models A and B simulations (D[Pα-C3'] =

4.6 Å and D[Pα-O3'] = 3.6 Å: Model A, and D[Pα-C3'] = 4.4 Å and D[Pα-O3'] = 3.5 Å: Model B) (see Table 5 and 2).

Table 6 compares the RMSD of the active-site amino acids (D110-Y115 and M184-D186) in four different active-site models with either D185 or D186 protonated for both Models A and B. The RMSDs for the protein backbone atoms and for all protein heavy atoms of three models (HD185-Model A, HD185-Model B and HD186-Model A) are slightly higher than the previously relative models A and B. It indicated that these models have little increasingly conformational changes of the active-site amino acids which might be either the $_1Mg^{2+}$ ligands or the $_2Mg^{2+}$ ligands, compared to the Models A and B. Whereas HD186-Model B showed the highest RMSD values corresponding to the noticeably conformational changes in polymerase active site as mentioned above.

Table 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.

| Model | RMSD (Å) | |
	Protein Backbone Atoms	All Protein Heavy Atoms
HD185-Model A	0.64	0.92
HD185-Model B	0.72	1.08
HD186-Model A	0.60	0.89
HD186-Model B	1.33	1.46

Taken together, the results from simulations of RT with one or the other aspartate residues (D185 and D186) protonated do not agree as well with the experimental structure. Therefore we conclude that their sidechains are likely to be present in the basic (carboxylate) form.

Combined quantum mechanical and molecular mechanical approach on the RT/dsDNA/dTTP ternary complex

1. QM/MM minimizations

From our results above, it appears that molecular dynamics simulations with these molecular mechanics (MM) methods can reproduce the structure of the polymerase active site from the crystal structure, and model the substrate complex of RT. However, standard MM is not a suitable tool to study the bond-forming and bond-breaking reactions. Our ultimate aim is to model the reaction mechanism of enzyme. We have therefore tested combined QM/MM approaches for this enzyme system.

In our QM/MM model, the QM region contains the reacting part (the methanol portion of deoxyribose ring of the 3'-terminal DNA primer and the methyltriphosphate of substrate complexed with two Mg^{2+} ions) and the coordination environments of the divalent cations, as well as the magnesium ions themselves. We used the same parameters and boundary restraints as treated in the MM stochastic boundary molecular dynamics (SBMD) calculations, except that no cutoff was applied for nonbonded interactions in the QM/MM calculations. The performance of two different semiempirical QM/MM methods (AM1/CHARMM and PM3/CHARMM) was tested by energy minimizations on Models A and B where details of the QM/MM energy minimization are given in Methods above. The starting structure for these QM/MM calculations was obtained from the MM minimized snapshot taken from the final MD structure. Table 7 shows the 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). The results strongly depend on the semiempirical methods used. The distance between the α-phosphorus of the dTTP substrate and the 3'-hydroxyl oxygen on the DNA primer terminus is important by reason of its bearing on the reaction coordinate of nucleotide addition. Compared to the Pα-O3' distance from the minimized MM snapshots at 500 ps of the simulation trajectories (3.60 Å: Model A and 3.45 Å: Model B), this distance

increased (to 3.81 Å: Model A and 4.18 Å: Model B) on AM1/CHARMM minimization. PM3/CHARMM predicted this distance to be also longer (4.47 Å: Model A and 3.81 Å: Model B). Due to its importance in the elimination of pyrophosphate (PPi), the Pα-O3α bond-breaking distance was also determined. Table 7 shows that for both methods, this distance increased (to 1.6 Å: AM1/CHARMM and 1.7 Å: PM3/CHARMM) from the minimized MM structures (1.5 Å) to be closer to that of the crystal structure (1.7 Å). Other distances between the α-phosphate oxygen atoms and the α-phosphorus are similar between these two methods, though AM1/CHARMM gives shorter distances (see Table 7). Both QM/MM methods generated reasonable conformations of the triphosphate moiety of the dTTP substrate in either its fully deprotonated or mono-protonated form (protonated on the Oγ2-oxygen), which agreed well with the crystal structure of dTTP, as shown in Figure 15(a) and 15(b). The binuclear magnesium coordination environment differs between these two semiempirical methods, as shown by the Mg-O distances in Table 7. AM1/CHARMM gives all Mg-O distances in average value close to 2.2 Å, while PM3/CHARMM predicts slightly shorter average distances (1.9 Å) except for the Mg1-O1α distance: Model A, and for the two distances (D[Mg1-O1α] and D[Mg1-O1γ]): Model B). With AM1/CHARMM, the Mg1-O1α distance decreased from the minimized MM strucutures (3.09 Å: Model A and 2.95 Å: Model B) (to 2.17 Å: Model A and 2.18 Å: Model B). PM3/CHARMM gave this distance increased (to 4.01 Å: Model A) but decreased (to 2.50 Å: Model B). In Model A, the PM3/CHARMM Mg1-O1α distance increased, this is probably owing to the effect of the Mg1-Mg2 repulsive interaction (D[Mg1-Mg2] = 4.46 Å) which the catalytic magnesium ion ($_1$Mg^{2+}) further displaced from the crystal structure (3.57 Å) as shown in Figure 16(a).

Table 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

Distance (Å)	X-ray	Model A			Model B		
		MM	AM1/ CHARMM	PM3/ CHARMM	MM	AM1/ CHARMM	PM3/ CHARMM
Pα(dTTP)-C3'(primer)	4.69	4.59	4.45	5.03	4.41	4.38	4.31
Pα(dTTP)-O3'(primer)	-	3.60	3.81	4.47	3.45	4.18	3.81
Pα(dTTP)-O1α(dTTP)	1.48	1.49	1.52	1.56	1.49	1.53	1.62
Pα(dTTP)-O2α(dTTP)	1.49	1.47	1.48	1.53	1.47	1.47	1.50
Pα(dTTP)-O3α(dTTP)	1.68	1.49	1.62	1.70	1.52	1.64	1.72
Mg1-Mg2	3.57	3.97	3.44	4.46	3.78	3.28	3.34
Mg1-O3'(primer)	-	2.04	2.21	1.89	2.06	2.22	1.91
Mg1-O1α(dTTP)	3.03	3.09	2.17	4.01	2.95	2.18	2.50
Mg1-OD2(D110)	2.68	1.80	2.15	1.86	1.82	2.22	1.87
Mg1-OD1(D185)	2.05	1.82	2.33	1.85	1.82	2.16	1.87
Mg1-OD1(D186)	3.87	1.78	2.11	1.84	1.77	2.14	1.86
Mg1-OH2(WAT)	-	1.96	2.24	1.88	1.92	2.22	1.88
Mg2-O1α(dTTP)	2.21	1.91	2.35	1.93	1.89	2.21	1.92
Mg2-O2β(dTTP)	2.44	1.88	2.25	1.90	1.89	2.23	1.88
Mg2-O1γ(dTTP)	2.39	1.84	2.19	1.91	1.89	2.28	2.52
Mg2-OD1(D110)	2.13	1.87	2.17	1.91	1.83	2.18	1.87
Mg2-O(V111)	2.27	2.02	2.22	1.92	1.96	2.18	1.91
Mg2-OD2(D185)	2.36	1.84	2.18	1.93	1.85	2.15	1.90

As can be seen in Figure 16, both QM/MM methods (AM1/CHARMM and PM3/CHARMM) showed some structural rearrangements of the coordination environment of the $_1Mg^{2+}$ ion because of the restored 3'-OH terminal primer, for both Model A and Model B. They shared this feature with their minimized MM (CHARMM) structure (see Table 7). In contrast, in the QM/MM optimized structures, the coordination number and arrangement of ligands observed around the nucleotide-binding magnesium ion ($_2Mg^{2+}$) and the conformation of the dTTP triphosphate are in very close agreement with those of the crystal structure. The two different semiempirical QM/MM methods preserve the basic features of the experimental structure, especially at the site of $_1Mg^{2+}$ ion. They provide structures similar to that of the crystallographically derived active site. These findings suggest that either of the two semiempirical QM methods, AM1 and PM3, may be suitable for this QM/MM modelling of the RT system. Nevertheless, we would note that, in particular for the AM1/CHARMM technique, the treatment of phosphorus should be carefully monitored in any reaction modelling due to AM1 errors of distorted geometry in phosphorus-containing compounds as found in previous studies (Zurek *et al.*, 2004; Lopez *et al.*, 2003).
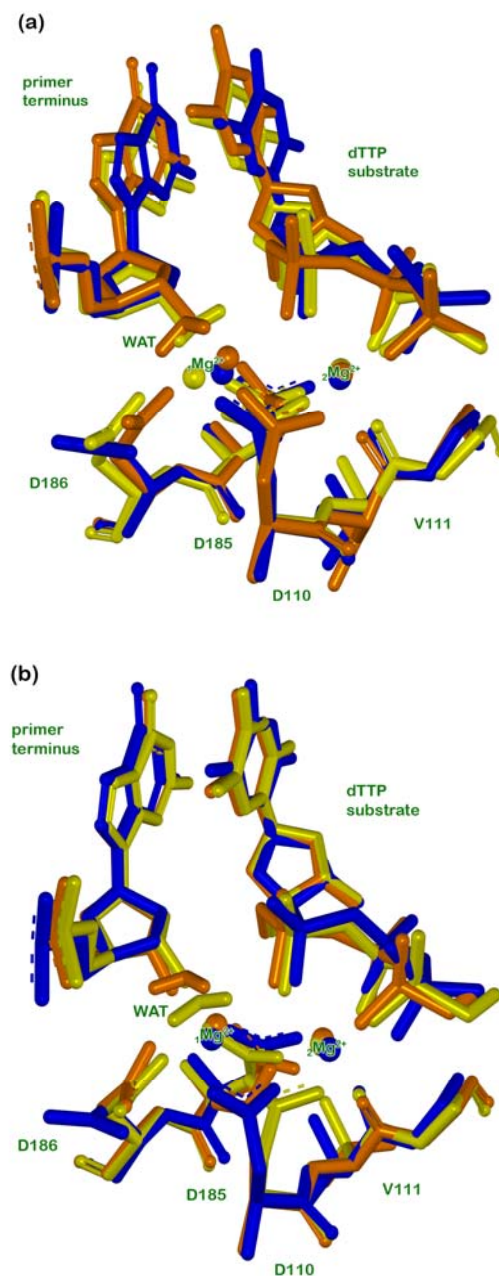
Figure 16   Superimposition of QM/MM minimized AM1/CHARMM (orange) and PM3/CHARMM (yellow) structures (from the final MD structure) compared to the crystal structure (blue). Model A is shown in 16(a) and Model B in 16(b). Hydrogen atoms are not shown except for the mono-protonated O2γ atom (Model B) and the bound water molecules.

Source: Rungrotmongkol (2006)
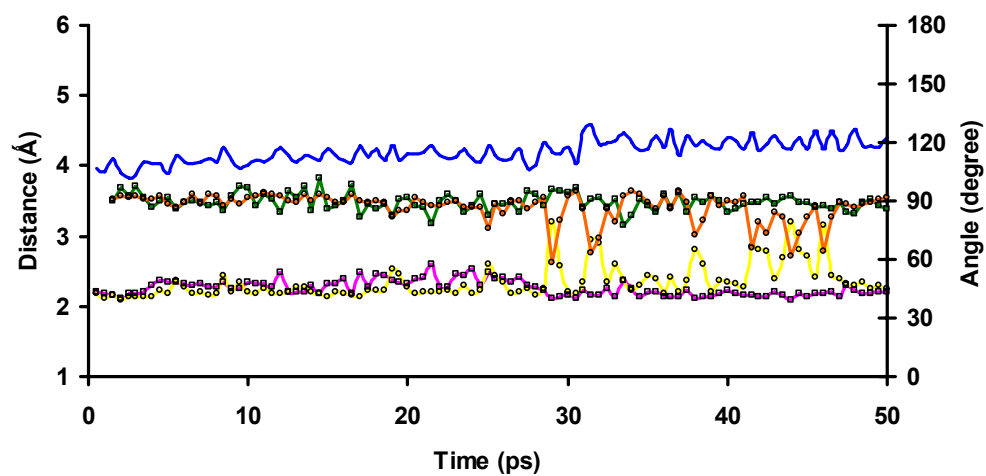
2. QM/MM dynamics simulations

To observe the dynamic properties of the active-site region involved the proton transfer reaction to a suitable base, thus, the reacting distances between H3T-hydrogen atom on the 3'-hydroxyl group of primer terminus and any possible base oxygen located in the 3'-OH pocket are determined. Based on two different semiempirical QM/MM dynamic simulations (AM1/CHARMM and PM3/CHARMM) for 50 ps, we found that three oxygen atoms have their positions closed to the 3'-OH terminal primer: these are the OD1-oxygen of two aspartates (Asp185 and Asp186), and the O5'-oxygen of the dTTP substrate. The plots of distance between the H3T-hydrogen and the proton acceptor and of angle of the O3'-proton donor---H3T---OD1/O5'-proton acceptor are shown in Figure 17.

The dynamic properties are strongly depends on each semiempirical QM/MM used. In fully deprotonated triphosphate system (Model A), AM1/CHARMM gives a hydrogen bonding interaction between the 3'-OH group of terminal primer and the carboxylate group of Asp186 in the first 12 ps simulations. Afterward, the H3T-hydrogen changes its position to sit in between two aspartate side chains (Asp185 and Asp186) which is also found in the mono-protonated system (Model B) suggesting that their OD1-oxygen are possible to abstract the proton from the 3'-OH group. In contrast, PM3/CHARMM shows the presence of a strong hydrogen bond between this 3'-OH moiety and the OD1-carboxylate oxygen of Asp185 in Model A or of Asp186 in Model B. This feature indicates that Asp185 or Asp186 might serve as a general base in reaction of deprotonation catalyzed by HIV-1 RT prior to subsequent nucleophile addition. However, another viable mechanism could involve the transfer reaction of the proton from the 3'-OH group to one of nonbridging oxygens of the α-phosphate which the most suitable oxygen might be the O5'-oxygen due to its orientation toward the 3'-OH group and forming a highly firmed hydrogen bonding interaction as observed in the classical molecular dynamic simulations.
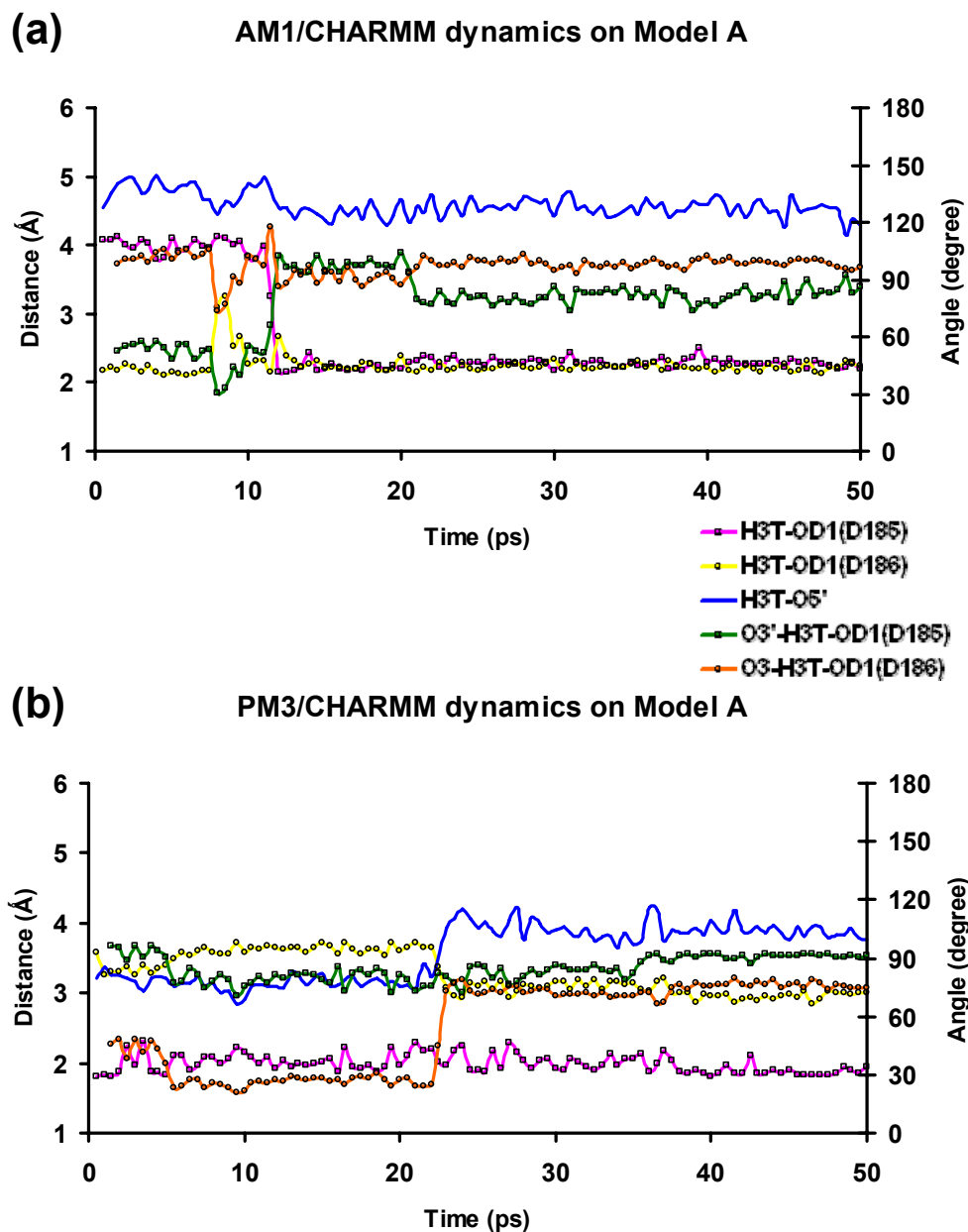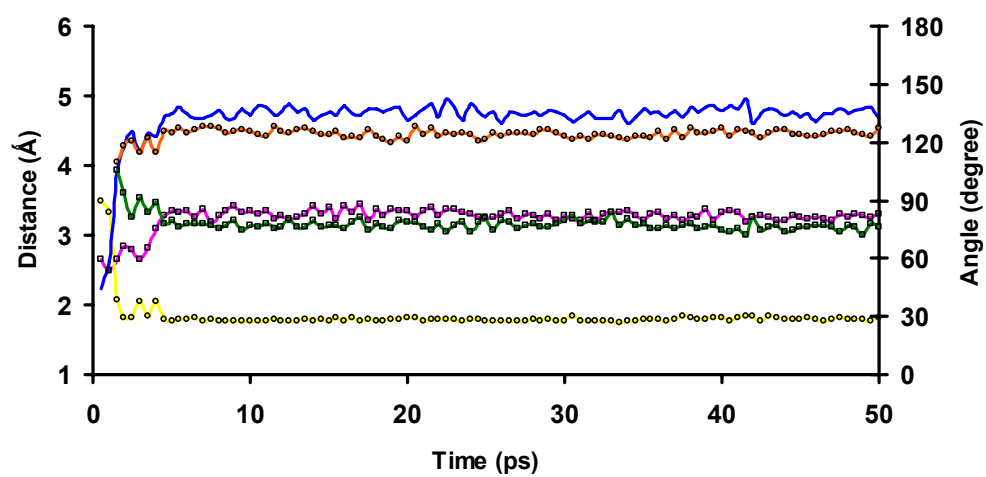
Figure 17   Plot of selected distances and angles as a function of time from two different semiempirical QM/MM dynamics simulations for 50 ps: (a) AM1/CHARMM on Model A; (b) PM3/CHARMM on Model A; (c) AM1/CHARMM on Model B; (d) PM3/CHARMM on Model B.

(c) **AM1/CHARMM dynamics on on Model B**


(d) **PM3/CHARMM dynamics on Model B**

Figure 17   (cont'd)

3. QM/MM approach in conjunction with umbrella sampling method

3.1 Mechanistic investigation in the small QM/MM partition scheme

Before beginning umbrella sampling reaction simulations, a QM/MM molecular dynamics simulation was performed with semiempirical (AM1) QM/MM minimization then MD simulations for 10 ps. We observe some rearrangements in the magnesium coordinate spheres. The $_1Mg^{2+}$ loses two water molecule ligands becoming its second nonbridging oxygens while Asp185 group is compensated. The $_2Mg^{2+}$ coordination in the result shows the lost of Asp110 group. After deprotonation, the protonated carboxyl oxygen of Asp185 moves away from $_1Mg^{2+}$ and $_2Mg^{2+}$ ions, so both magnesium coordinations remain only four ligands. The proton-transfer reaction from 3'-OH group of primer to carboxyl oxygen of Asp185 is nonspontaneous in aqueous solutions. The calculated free energy barrier (*ca.* 45 kcal/mol) is very high as shown in Figure 18. It is unrealistically high: this is probably because the AM1 semiempirical method overestimates the activation energy. AM1 suffers from some known limitations, e.g. in the treatment of phosphorus.
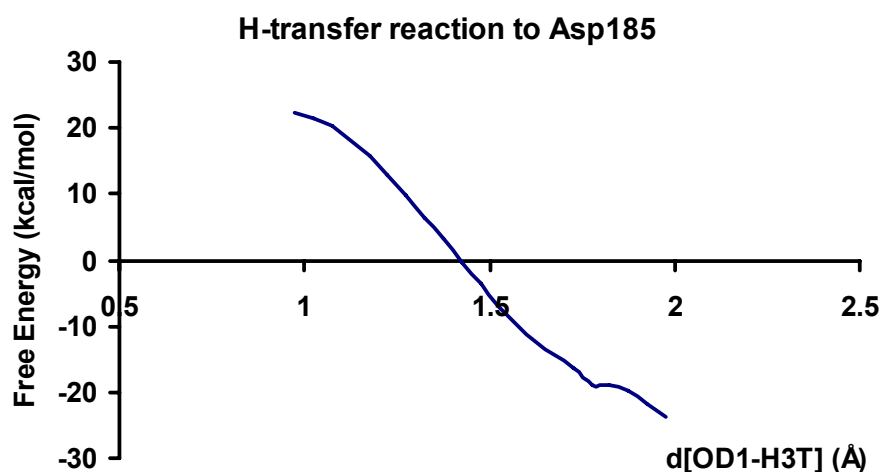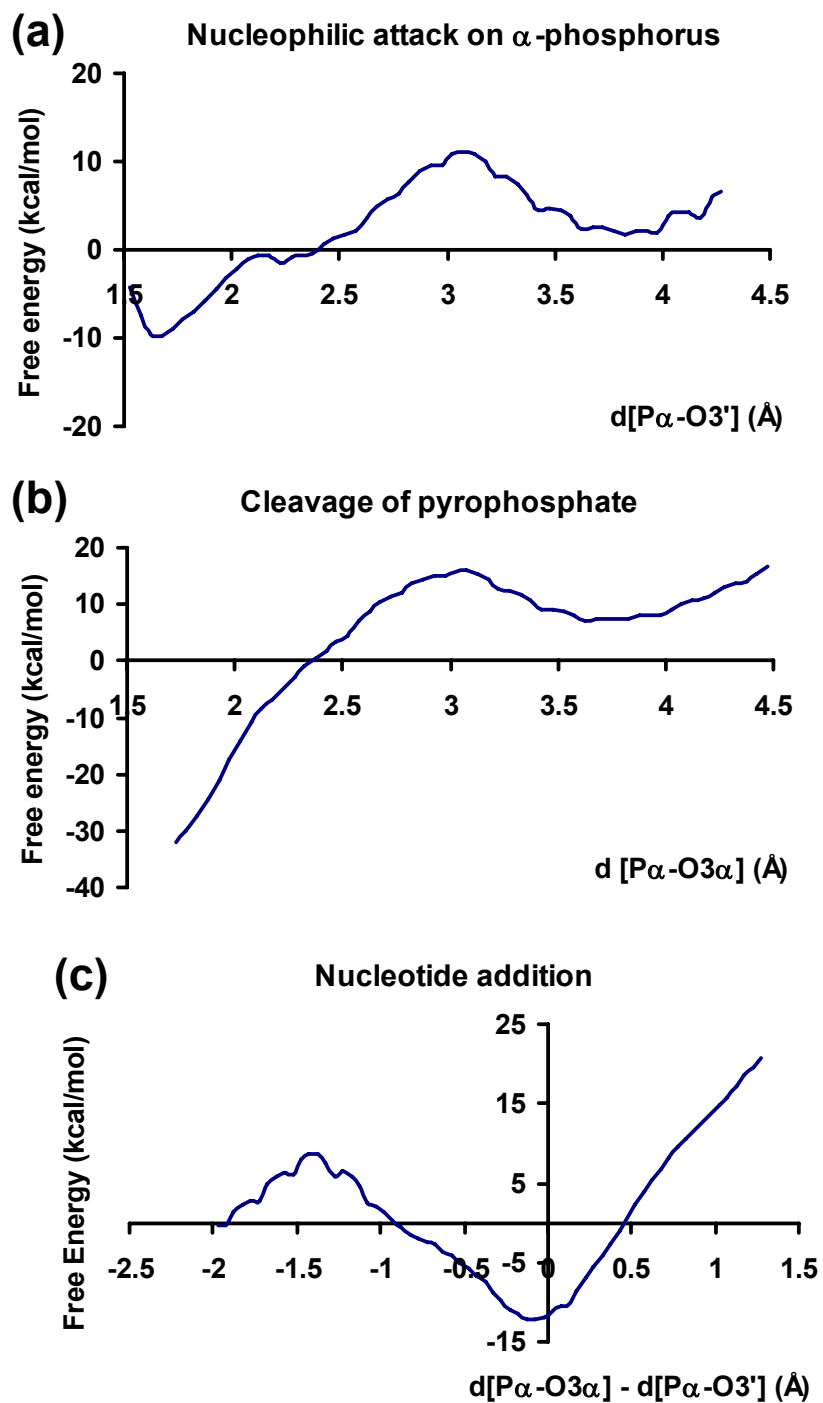

Figure 18 Free energy profiles for reaction coordinate as hydrogen primer terminus-carboxyl oxygen (OD1) of Asp185.

Source: Rungrotmongkol (2004)

In the nucleophilic addition, the Pα atom of methyl triphosphate $(CH_3P_3O_{10})^{4-}$ forms a new chemical bond with the O3'-oxygen atom of deoxyribose ring in its deprotonated state and the Pα-O3α bond is broken to gain the phosphate diester and pyrophosphate as final products. Based on two reaction coordinates (Figure 19), the energy profiles for nucleophilic attack are *ca.* 10 kcal/mol and the pentacovalent phosphorus intermediate is formed prior to the departure of the leaving group, pyrophosphate. Hence, the results support a stepwise mechanism, proceeding via a relatively stable pentacoordinate intermediate. The structure of the pentacoordinate intermediate (bonded to 5 oxygen atoms) is shown in Figure 20. Some important structural parameters involving the α-phosphorus and oxygen atoms are compared in Table 8. While the ribose ring of primer moves to triphosphate moiety of dTTP substrate, decreasing Pα-O3' distance, all Pα and oxygen bonds are longer, especially, Pα-O3α breaking bond. In known structures of pentaoxyphosphoranes, the P–O distances fall in the range 1.57–1.76 Å.23 This distance could still be longer, depending on the substituents and on the extent to which there is contribution from coordinate covalent character (Chandrasekaran *et al.*, 2003; Swamy *et al.*, 2003). The intermediate is further stabilized by hydrogen bonding interactions between the triphosphate moiety and the enzyme backbone of Asp113 and Ala114, the guanidinium group of Arg74 and the ε-amino group Lys65 corresponding in the crystal structure. These hydrogen bonding interactions are shown in all simulations but only change the triphosphate oxygen atoms.

(a)   **Nucleophilic attack on α-phosphorus**

Free energy (kcal/mol)

d[Pα-O3'] (Å)

(b)   **Cleavage of pyrophosphate**

Free energy (kcal/mol)

d [Pα-O3α] (Å)

(c)   **Nucleotide addition**

Free Energy (kcal/mol)

d[Pα-O3α] - d[Pα-O3'] (Å)

Figure 19   Modeled structure represented for the active site. QM region is shown
shaded with four hydrogen link atoms cycled.

Source: Rungrotmongkol (2004)

Table 8 The average structural parameters between α-phosphorus and oxygen or
carbon in the nucleophile addition.

Distance (Å)	Product1*	Intermediate	Product2
Pα-C3'(primer terminus)	3.91	2.73	2.52
Pα-O3'(primer terminus)	3.72	1.70	1.59
Pα-O3α(dTTP)	1.59	1.68	3.69
Pα-O1α(dTTP)	1.54	1.59	1.52
Pα-O2α(dTTP)	1.46	1.52	1.48
Pα-O5'(dTTP)	1.59	1.60	1.60

*Product1 is obtained from deprotonation and product2 is the final product.
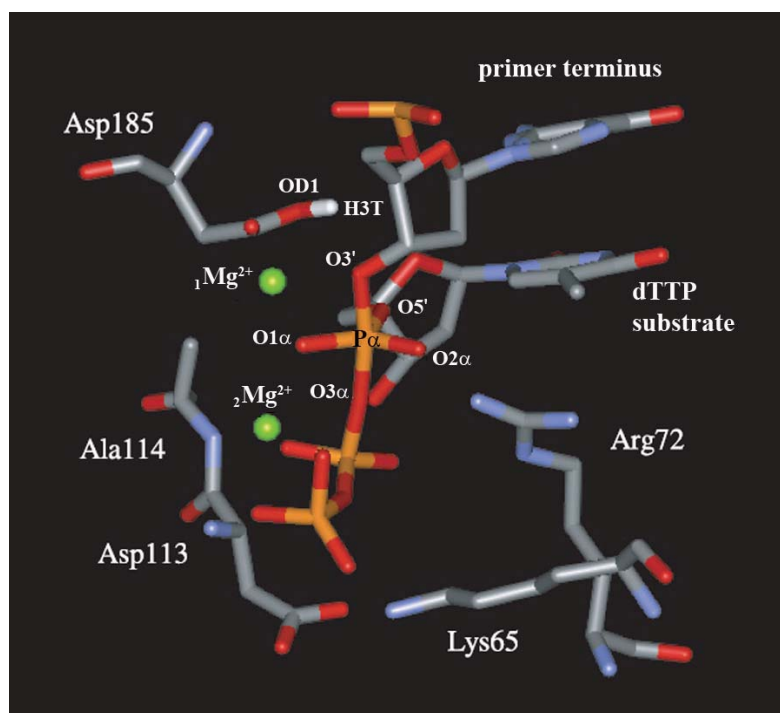
Figure 20   Pentacovalent intermediate, magnesium coordination environment and
possible hydrogen bonds taken from AM1 QM/MM simulation snapshot.
H atoms bonded to the C and N atoms are not shown.

Source: Rungrotmongkol (2004)

3.2 Mechanistic investigation in the large QM/MM partition scheme

At the last MM MD snapshot of deprotonated triphosphate system (Model A), we have performed PM3/CHARMM minimization and MD simulation for 50 ps. Starting from this equilibrated model, the system was applied by PM3/CHARMM approach in conjunction with umbrella sampling for 1 ps simulation and following by 29 ps simulation to sample the reaction coordinates for the H-transfer reaction with three different H-acceptors (Asp185, Asp186 and dTTP) and then for the nucleotide addition. The latter step is comprised of nucleophilic attack from the negative charged O3'-hydroxyl group of the terminal DNA primer to α-phosphate of dTTP substrate leading to the formation of pentacoordinated structure. We expect that this structure should be a stable intermediate as found in the small QM/MM partition scheme. Subsequently, the Pα-O3α bond is broken to gain the newly formed 3'-5' phosphodiester and pyrophosphate as final product complex. Another objective is to observe the magnesium coordination environment and H-bond interactions during the reaction pathway from the reactant to product complexes.

As approximated reaction coordinate for both mechanistic reactions, defined as the difference between the breaking and making bonds: these are d[O3'(primer)-H3T(primer)] − d[OD1(base)-H3T(primer)] for deprotonation and d[O3α(dTTP)-Pα(dTTP)] − d[O3'(primer)-Pα(dTTP)] for DNA polymerization. With restrained dynamics, the reaction coordinate was sampled every 0.1 Å for 30 ps QM/MM MD simulations. The force constants are equal to 200 and 300 kcal/mol in deprotonation and DNA polymerization, respectively.

3.2.1 Deprotonated triphosphate system (Model A)

The free energy profiles obtained on the basis of these simulations for three different reaction pathways in deprotonation (H3T-hydrogen atom transfers to Asp185, Asp186 and dTTP) from reactant to product complexes are presented in Figure 21. Three different profiles, with distinguishable energy barriers, were obtained when different H-acceptor involved in reaction. The reactant complex (Figure 22) of deprotonated triphosphate system (Model A) shows that the $_1Mg^{2+}$ ion is in the position where it can polarize the 3′-OH group of the primer strand facilitating deprotonation of the 3′-OH group and leading to the creation of the attacking nucleophile on α-phosphorus atom. The H-transfer reaction from the terminal primer to the OD1-carboxylate oxygen of two aspartates (Asp185 and Asp186) and the O5'-oxygen of dTTP, consume energy c.a. 15.0 (Asp185), 17.9 (Asp186) and 23.8 (dTTP) kcal/mol. With the lowest energy barrier for deprotonation found in Asp185 as a base mechanism, thus, Asp185 may be the most suitable hydrogen acceptor to abstract the H3T-hydrogen atom from the 3'-OH primer terminus. The proton is now attached to Asp185 and this aspartic acid has a hydrogen bonding interaction toward the negative charged O3'-hydroxyl oxygen on primer terminus which the average H3T-O3' distance is 1.62 Å (SD = 0.01 Å) and average OD1-H3T-O3' angle is 148.5 degree (SD = 2.1 degree) as shown in Table 9 (see also Figure 23).

Whereas the H-transfer reaction via another aspartate (Asp186) mechanism has used the energy barrier closed to 17.9 kcal/mol gaining the aspartic acid and the deprotonated deoxyribose ring at the primer terminus. Although, this product complex (11.1 kcal/mol) seems to be more stable in relative with the product complex (13.6 kcal/mol) from the previous reaction pathway (Asp185 as a proton acceptor), its distance between this free O3'-hydroxyl oxygen and α-phosphorus on deprotonated triphosphate, importance in consequent reaction coordinate for nucleophilic attack reaction, is increased (from 3.58 Å) to be 5.51 Å. From this point of view, it is unlikely because both the primer terminus and the dTTP triphosphate move apart

from each other resulting unsuitable orientation to happen the reaction of nucleophile addition.

For the proton transfer reaction to the O5'-oxygen of dTTP substrate, the reaction barrier was increased (to be c.a. 23.8 kcal/mol) compared to those reactions involved an aspartate leading to unstable product with high energy c.a. 24.2 kcal/mol. These results indicated that Asp186 and dTTP are unlikely to act as H-acceptor relatively compared to Asp185 in either higher energy barrier for H-transfer reaction or long attacking nucleophile distance, probable leading the enlarged overall energy barrier. Thus, we exclude the mechanistic investigation of following nucleophile addition after the proton transfer reaction to either Asp186 or dTTP in the fully deprotonated model. It is a worth to note that Asp185 plays an important role as a catalytic base and accepts the proton from the 3'-OH terminal primer in this modelled system. Thus, only the product complex from the H-transfer reaction to Asp185 is used as starting structure for further investigation on the DNA polymerization.
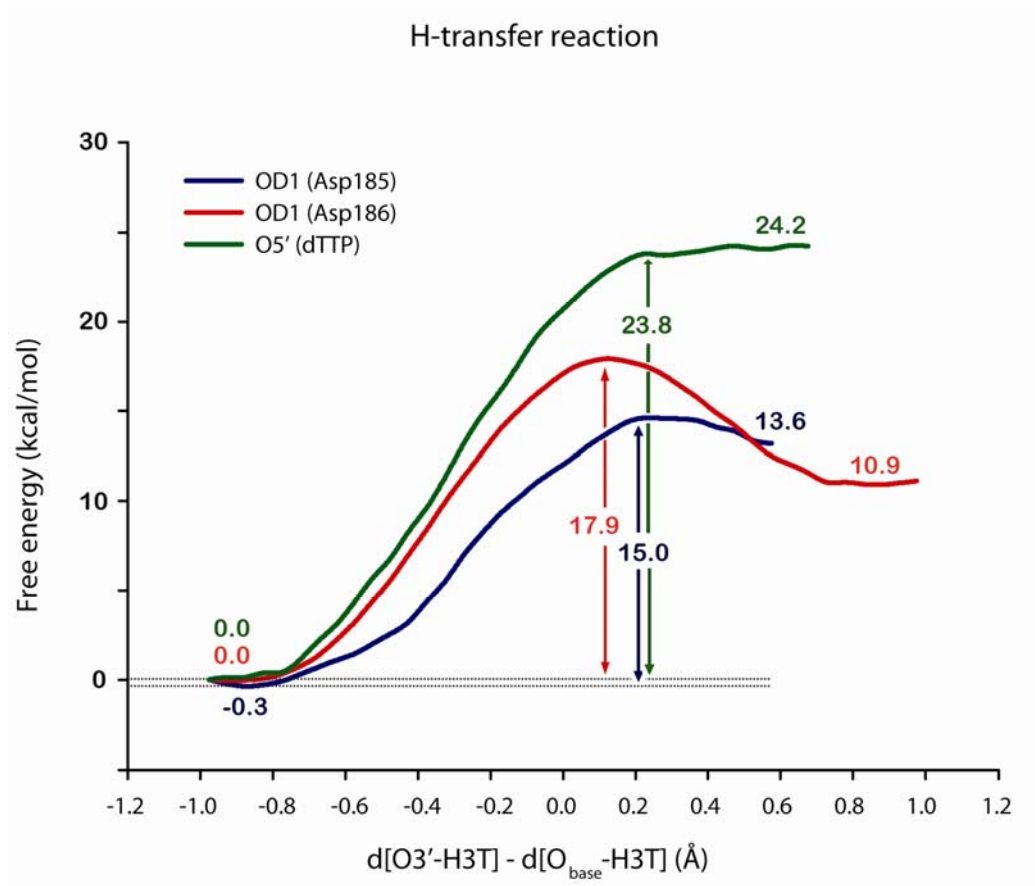
Figure 21   Free energy profiles for three different H-transfer reactions in Model A from H3T-hydrogen on the 3'-OH primer terminus to a H-accepter (Asp185, Asp186 or dTTP) using reaction coordinate as different reaction coordinate between breaking and forming bond.
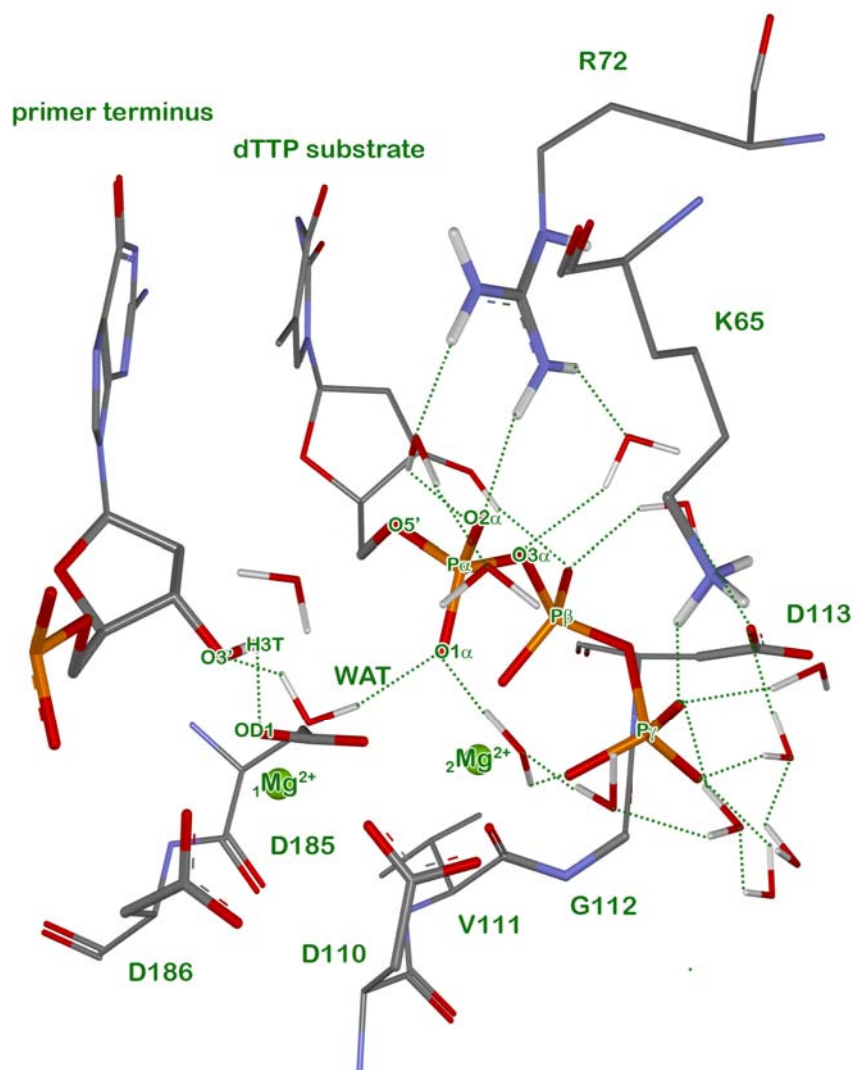
Figure 22   Reactant complex (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.
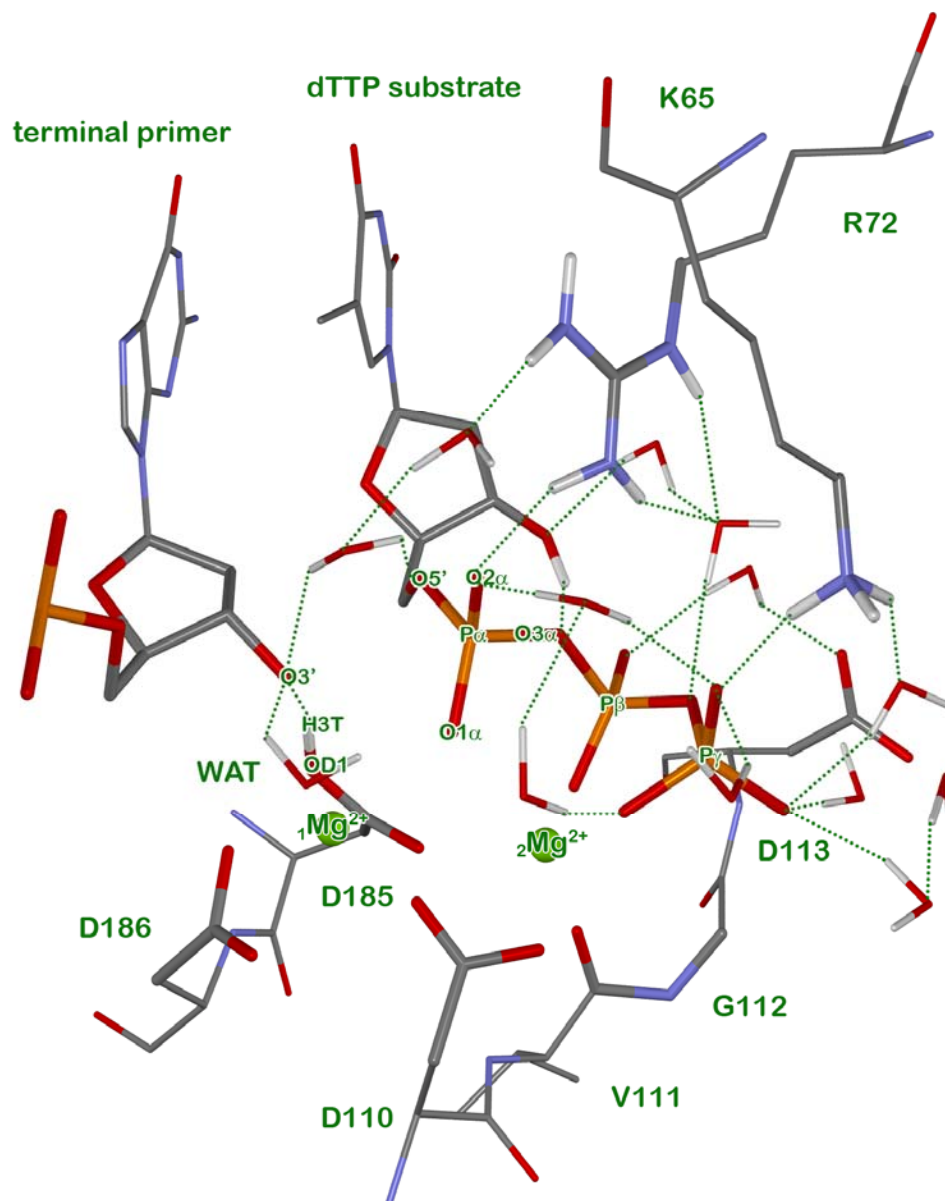
Figure 23   Product complex (Model A) from H-transfer reaction via Asp185, magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.

Table 9 Some parameter (Å) of critical structures in mechanistic reaction of HIV-1 RT via D185 as normal base

|	Reactant	Product	Intermediate	Final Product
Mg1 coordination sphere:				
D[Mg1-O3'(primer)]	2.49 (0.01)	2.45 (0.01)	2.48 (0.01)	2.50 (0.00)
D[Mg1-O1α(dTTP)]	2.52 (0.01)	2.50 (0.01)	2.42 (0.01)	2.43 (0.00)
D[Mg1-OD2(D110)]	1.87 (0.00)	1.85 (0.01)	1.85 (0.01)	1.85 (0.00)
D[Mg1-OD1(D185)]	2.46 (0.00)	2.45 (0.00)	2.45 (0.00)	2.45 (0.00)
D[Mg1-OD1(D186)]	1.85 (0.00)	1.84 (0.01)	1.83 (0.00)	1.83 (0.00)
D[Mg1-WAT5884]	1.88 (0.01)	1.86 (0.00)	1.86 (0.01)	1.86 (0.01)
Mg2 coordination sphere:				
D[Mg2-O1α]	2.49 (0.01)	2.47 (0.01)	2.46 (0.01)	2.47 (0.00)
D[Mg2-O2β]	1.86 (0.01)	1.84 (0.01)	1.85 (0.00)	1.85 (0.00)
D[Mg2-O1γ]	1.86 (0.00)	1.86 (0.01)	1.86 (0.00)	1.86 (0.00)
D[Mg2-OD1(D110)]	1.86 (0.01)	1.85 (0.00)	1.86 (0.01)	1.86 (0.00)
D[Mg2-O(V111)]	2.47 (0.01)	2.48 (0.02)	2.47 (0.01)	2.49 (0.01)
D[Mg2-OD2(D185)]	2.50 (0.01)	2.48 (0.01)	2.48 (0.01)	2.50 (0.01)
D[Mg1-Mg2]	3.41 (0.03)	3.67 (0.05)	3.65 (0.02)	3.68 (0.01)
D[H3T(primer)-O3'(primer)	0.92 (0.03)	1.62 (0.01)	1.81 (0.08)	1.85 (0.06)
D[H3T(primer)-OD1(D185)	1.91 (0.02)	1.01 (0.00)	0.98 (0.00)	0.97 (0.00)
θ[O3'-H3T-OD1]	146.7 (5.0)	148.5 (2.1)	150.6 (3.7)	147.4 (2.4)
D[Pα(dTTP)-C3'(primer)]	4.39 (0.10)	4.25 (0.12)	2.92 (0.02)	2.79 (0.01)
D[Pα(dTTP)-O3'(primer)	3.58 (0.08)	3.33 (0.12)	1.89 (0.01)	1.76 (0.00)
D[Pα(dTTP)-O3α(dTTP)	1.70 (0.01)	1.68 (0.03)	1.88 (0.01)	4.36 (0.01)
θ[O3'-Pα-O3α]	151.3 (5.6)	147.3 (4.7)	157.9 (3.1)	148.2 (1.6)
φ[Pα-O5'-O2α-O1α]	-24.1 (0.7)	-22.1 (0.8)	-0.6 (0.6)	20.5 (0.5)

Next step, to investigate the nucleotide addition after the terminal primer ribose ring was activated by losing its proton (H3T-hydrogen atom), the reaction coordinate was defined as different distance between $P\alpha$-$O3\alpha$ breaking and $P\alpha$-$O3'$ forming bond starting from -1.6 Å to 3.0 Å. The free energy profile for this reaction was presented in Figure 24. Starting structure obtained from the product complex of previous H-transfer reaction to Asp185, the negative charged $O3'$ hydroxyl atom moves closer to attack on α-phosphorus of the deprotonated dTTP by proceeding with 2.3 kcal/mol activation barrier which results in the formation of the pentacoordinated structure, α-phosphorus atom bonded with five oxygen atoms (see Figure 25), located at reaction coordinate approximately equal to 0 Å. This structure is an intermediate as also found in the small QM/MM partition scheme. It was found that the pentacovalent intermediate is energetically closed to 13.5 kcal/mol in comparable to the starting structure of this reaction pathway (13.6 kcal/mol). In petacovalent intermediate, the negative charge of triphosphate is mainly contributed on the β- and γ-phosphate moieties which increasingly stabilized by hydrogen bonding interactions with its hydroxyl group on the deoxyribose ring, the methylammonium of Lys65, the guanidinium of Arg72, the backbone (NH) of Asp113 and some water networks (Figure 25). The last reaction step is $P\alpha$-$O3\alpha$ departure taken an additional energy approximately 4.9 kcal/mol leading to the final product complex (the newly formed $3'$-$5'$ phosphodiester and pyrophosphate) with one DNA unit elongation of the primer chain (Figure 26). The free energy profile indicated that this final product complex is much more stable than both starting structure and intermediate complexes about 7 kcal/mol (see Figure 24). When the starting structure complex is driven to the product complex in the reaction of DNA polymerization, the average $P\alpha$-$C3'$ distance is significantly shorter (from 4.25 Å to 2.79 Å) relative to that of crystal structure (4.7 Å), consequently the average $P\alpha$-$O3'$ distance decreased (from 3.33 Å) to be 1.76 Å found in the $3'$-$5'$ phosphodiester. The inversion of configuration on the α-phosphorus atom is observed from change in the $P\alpha$-$O5'$-$O2\alpha$-$O1\alpha$ dihedral angle (H-transfer product: -22.1 degree; pentacovalent intermediate: -0.6 degree and final product: 20.5 degree) indicating the S_N2 reaction of DNA polymerization. The observation of a nearly perfect planar configuration characteristic of pentacoordinate intermediate is clearly seen in our calculation when the nucleophilic attacking distance (1.89 Å) is

almost equal to the Pα-O3α departure distance (1.88 Å). The inversion character of the reaction has been supported by the reactant and product geometries with stereochemical experiments (Burgers *et al.*, 1979; Brody *et al.*, 1982) and theorectical investigations (Abashkin *et al.*, 2001; Florian *et al.*, 2003). In the process of breaking down the Pα-O3α bond, an additional negative charge on the β-phosphate was stabilized by the nearby residue Arg72 whose side chain rotated to and formed stronger hydrogen bonding interaction with the broken O3α-oxygen atom. Furthermore, the pyrophosphate has the hydrogen bonding interactions with some tightly bound water molecules, the methylammonium of Lys65 and the backbone (NH) of Asp113 as can be seen from Figure 26. A hydrogen bonding interaction between the H3T-hydrogen on the general base (Asp185) and the O3'-hydroxyl oxygen of primer terminus was always observed during nucleotide addition along reaction coordinate from the starting structure to final product complexes as can be clearly seen in Figure 25 and 26 and also Table 9.

Taken into account, the energy barrier for overall reaction, H-transfer reaction and nucleotide addition, is closed to 18.4 kcal/mol while the H-transfer reaction to Asp185 becomes the rate-limiting step which needs 15.0 kcal/mol of activation energy. In addition, the relative energy of the final product complex is higher than the reactant complex about 6.5 kcal/mol (see Figure 22 and 24). In reverse reaction the reactant energy of the final product to the reactant complexes are approximately closed to 11.9, 2.4 and 1.4 kcal/mol, respectively as presented in Figure 22 and 24. These results show that the proposed reaction mechanism via Asp185 is the most feasible in this deprotonated triphosphate system.

In Asp185 as a proton acceptor mechanism, it is interesting to observe the magnesium coordination environment along the reaction pathway from reactant to final product complexes as shown in Table 9. It can be seen that both magnesium ions have same coordinating ligands along the reaction pathway from the reactant to product complexes. The catalytic $_1Mg^{2+}$ ion is nearly equal coordinated to three oxygen atoms concerning the reaction: these are the carboxyl OD1-oxygen on Asp185 side chain, the O3'-oxygen on deoxyribose ring, and the O1α-oxygen on α-phosphate,

with all three Mg1-O distances falling in the 2.4-2.5 Å range. Other its three oxygen ligands (the carboxyl OD2-oxygen on Asp110, the OD1- oxygen on Asp186 and the modelled water oxygen) are closer positioned to the $_1Mg^{2+}$ ion (in range of 1.8-1.9 Å). In the nucleotide-bridging $_2Mg^{2+}$ sphere, two ranges of Mg2-O distance are 2.4-2.5 Å and 1.8-1.9 Å leading to the presence of approximately octahedral shape similar to the $_1Mg^{2+}$ coordinating geometry. Thus, it can be explained in a straightforward manner that the divalent cations provide similar interactions to their coordinating ligands in overall reaction. In addition, the Mg1-Mg2 distance is slightly increased from the reactant to final product by 0.27 Å.
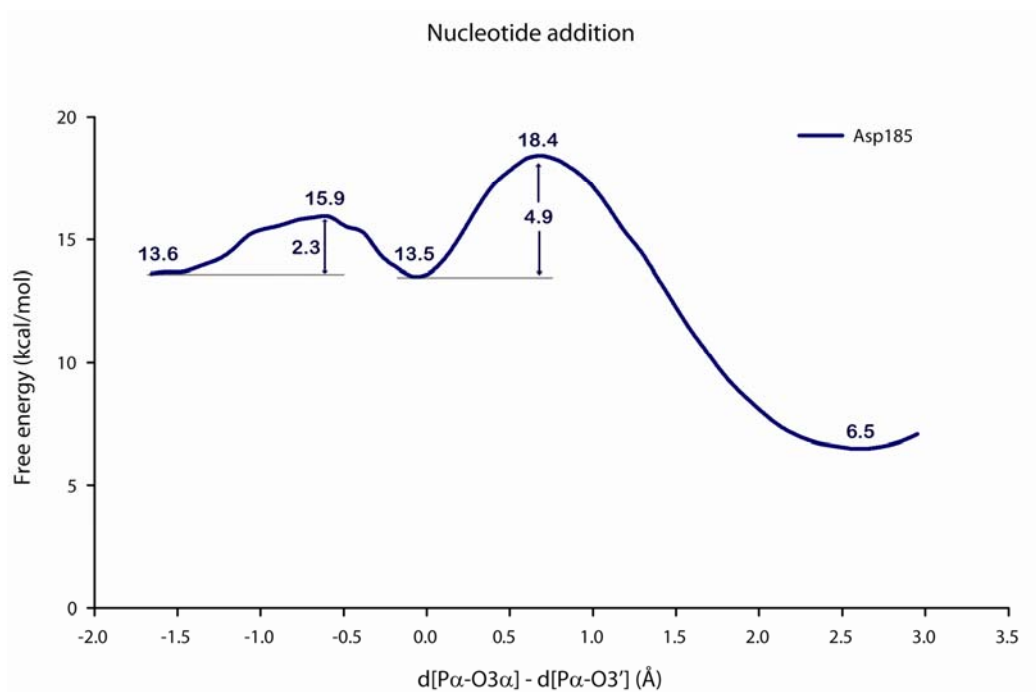
Figure 24   Free energy profiles for nucleophilic addition in reaction coordinate as: different reaction coordinate between Pα-O3α breaking and Pα-O3' forming bond.
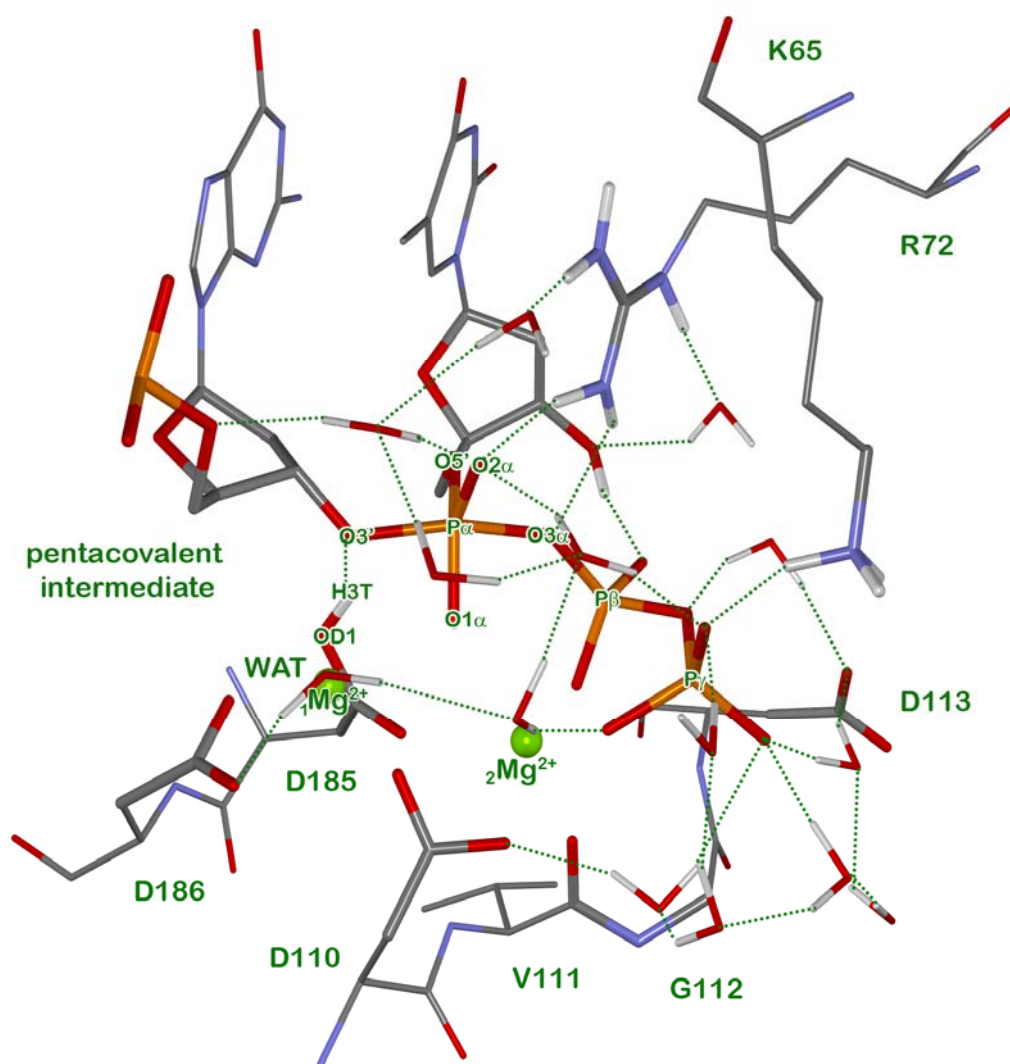
Figure 25   Pentacovalent intermediate (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.
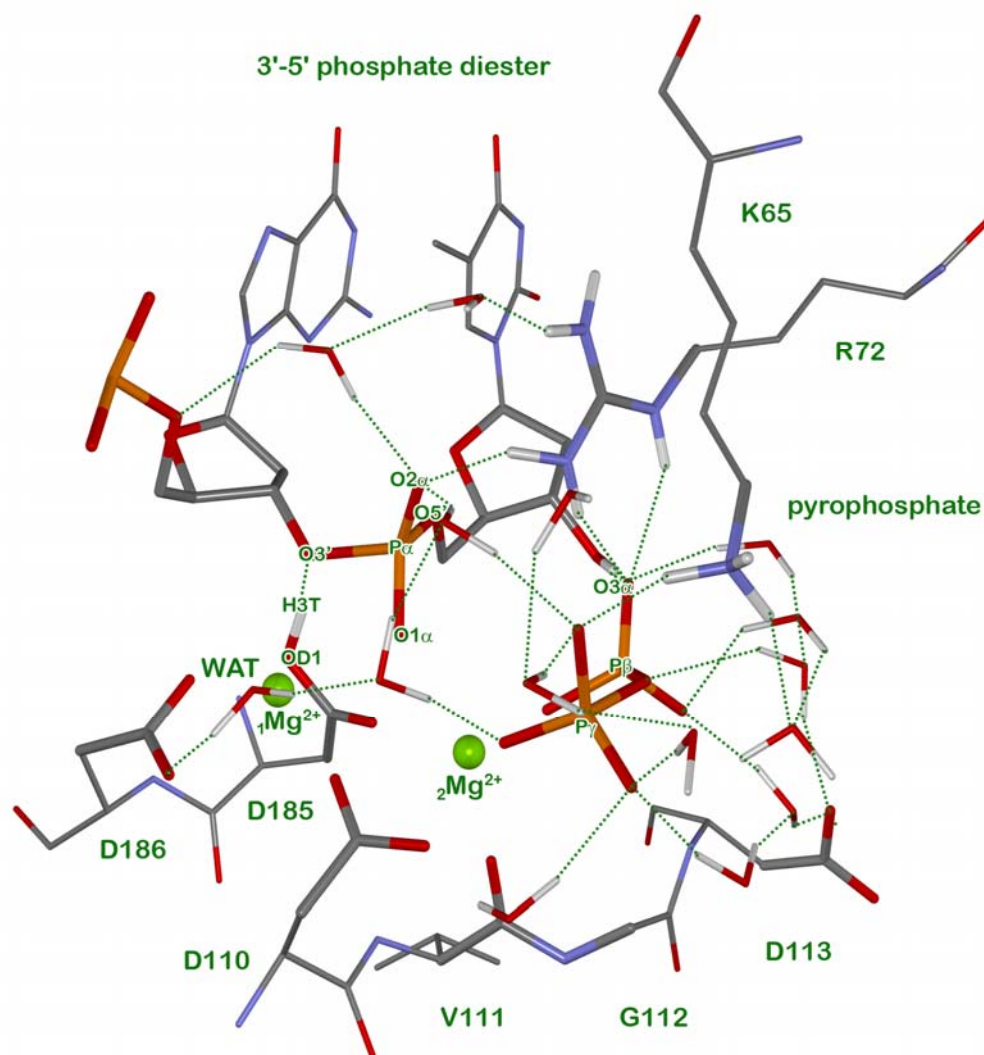
Figure 26   Final product complex (Model A), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.

### 3.2.2 Mono-protonated triphosphate system (Model B)

The free energy profiles for the reaction in the mono-protonated triphosphate system (Model B) were also calculated by combined semiempirical QM/MM MD simulations in conjunction with umbrella sampling. While the QM region consisting of 60 reacting atoms with the QM net charge of -2 was treated quantum mechanically by PM3 method, the reminder of system was considered classically by molecular mechanics (CHARMM). Figure 27 shows the free energy barriers for three different mechanisms in the H-transfer reaction from the 3'-hydroxyl group of the primer terminus. Prior to consider the free energy profiles, it may be worth to compare the reaction complex between two modelled systems (deprotonated form (Model A) and mono-protonated form (Model B) as given in Figure 22 and 28 respectively. A significant difference between these two reactant complexes, is that the deprotonated triphosphate dTTP in the reactant structure from Model A has the direct hydrogen bonding interaction with the guanidinium of Arg72 (O2α---NH2) while in Model B this side chain moves out from the triphosphate moiety. In latter, we found the indirect hydrogen bonding networks from the Arg72 guanidinium to the dTTP triphosphate moiety via two water molecules (see Figure 28).

In all three activation pathways, the lowest energy barrier was found in the Asp186 as a base mechanism approximately closed to 13.9 kcal/mol resulting in the energetic product at c.a. 11.2 kcal/mol. In comparison between two different protonation systems of the dTTP triphosphate, there are two main differences in the results for the H-transfer reaction to Asp186 mechanism. Firstly, the activation barrier in Model B is less than in Model A by 4 kcal/mol although the products are in the similar energy (c.a. 11 kcal/mol) relative to their own reactant complexes. Secondary, only product complex obtained from Model B shows the active conformation for the next reaction of nucleotide addition which the average attacking Pα-O3' distance is equal to 4.2 Å whereas this distance in Model A is longer (5.5 Å) as mentioned above.

Based on the H-transfer reaction involved Asp185 as a base mechanism, the results show the similarity of the activation barriers of about 15 kcal/mol in both models A and B but the product complex in Model A simulations has more slightly stability than an another by 2 kcal/mol compared to its reactant complex. According to the average $P\alpha$-O3' distance in the product complex is c.a. 3.7 Å. In structural analysis, we found the movement of the guanidinium of Arg72 to be little closer to the mono-protonated triphosphate moiety in the product complex resulting in the indirect hydrogen bonding interactions to the $O2\alpha$- and $O1\beta$-oxygen atoms on the triphosphate group.

The reaction free energy for H-transfer reaction from the 3'-OH deoxyribosyl group to the O5'-oxygen on the α-phosphosphate in the HIV-1 RT active site is estimated to be c.a. 23.4 kcal/mol which is highest activation barrier similarly also found in the previous system (Model A). It is, thus, expected that the overall reaction barrier may become even higher than the base mechanism via an aspartate (either Asp185 or Asp186 in the case of Model B). Consequently, only the aspartate as a base mechanism would be continuously investigated for the reaction of the nucleotide addition.
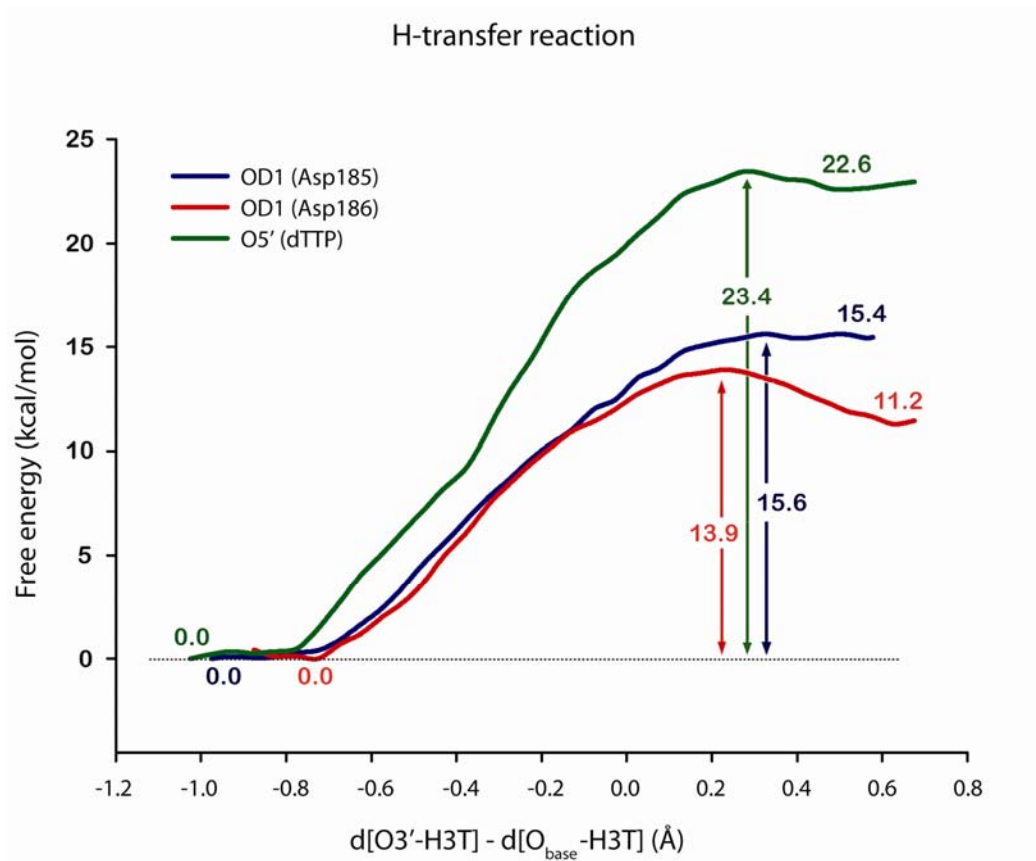
Figure 27   Free energy profiles for three different H-transfer reactions in Model B from H3T-hydrogen on the 3'-OH primer terminus to a H-accepter (Asp185, Asp186 or dTTP) using reaction coordinate as different reaction coordinate between breaking and forming bond.
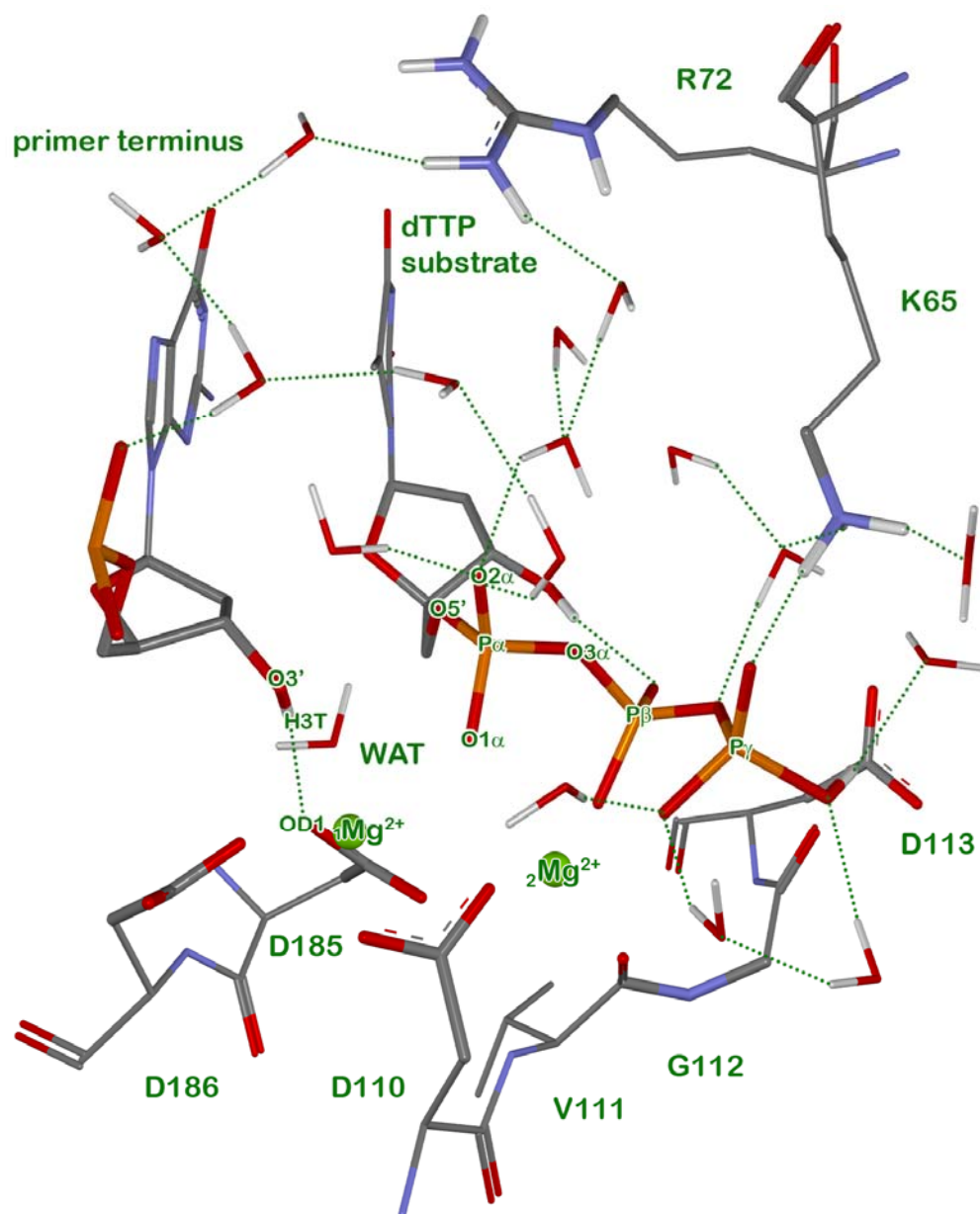
Figure 28   Reactant complex (Model B), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.
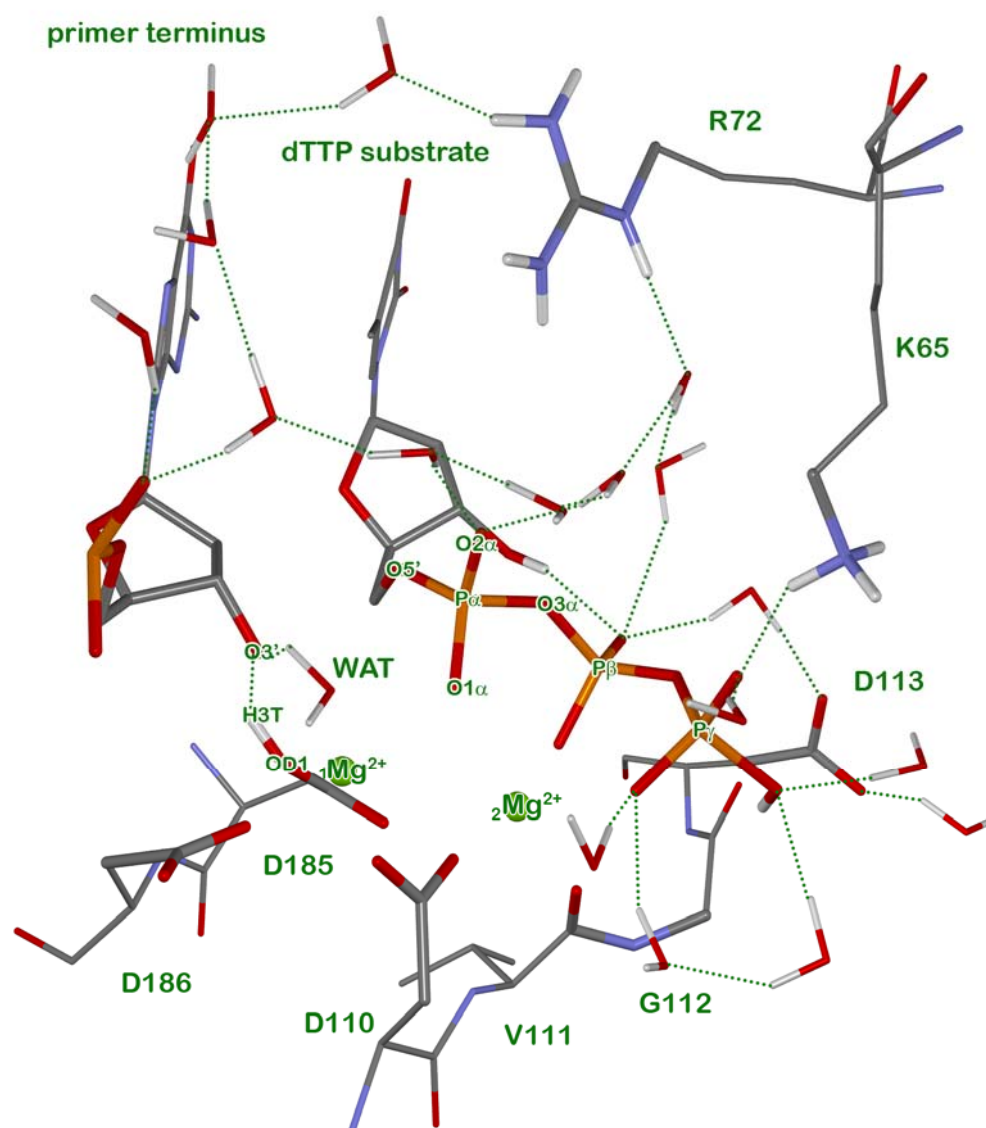
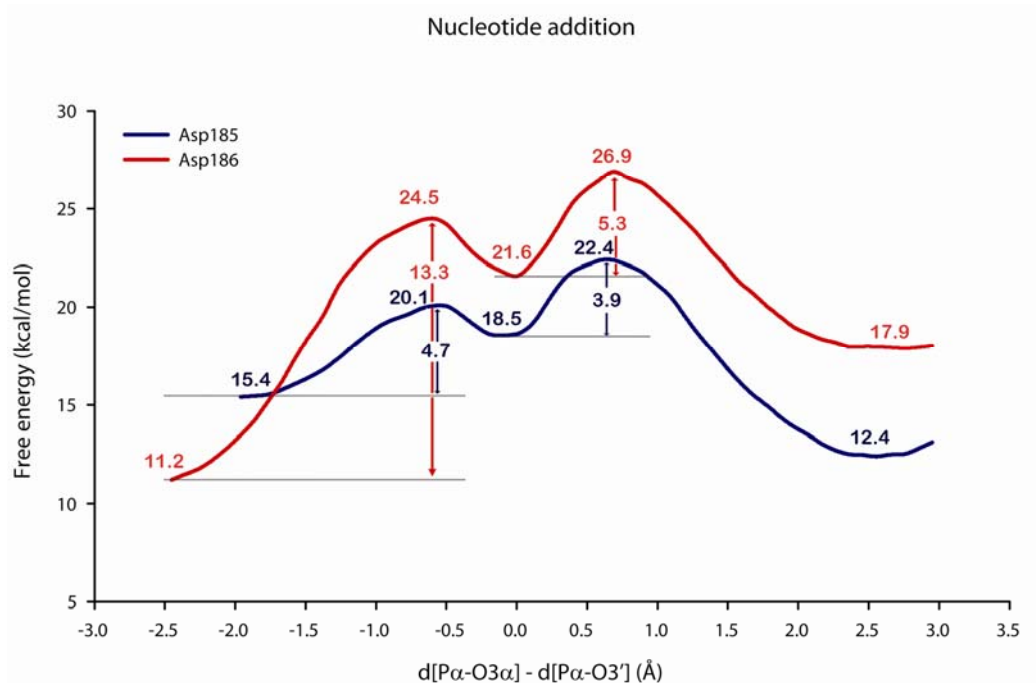**Figure 29**   Product complex (Model B) from H-transfer reaction via Asp185, magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules and on triphosphate moiety are shown.

In the investigation on the reaction of nucleotide addition, the starting structure was obtained from the product complex at the last reaction coordinate of deprotonation. Therefore, the starting values for the reaction coordinate defined as the difference between the Pα-O3' forming and the Pα-O3α breaking distances in the Asp185 and Asp186 as a base mechanism are -2.0 Å and -2.5 Å, respectively. The free energy profiles for the consequent nucleotide addition following these two different H-transfer reactions are shown in Figure 30. In Asp185 as a base mechanism, the activation barrier for the attacking on the α-phosphorus atom of triphosphate group is energetically closed to 4.7 kcal/mol leading to the formation of the pentacovalent intermediate complex, then the cleavage of pyrophosphate (PPi) taken place by consuming more 3.9 kcal/mol in the energy barrier as presented in Figure 31. In the last reaction, we observed the final product complex consisting of the newly-formed 3'-5' phosphate diester and the mono-protonated pyrophosphate at reaction coordinate closed to 2.5 Å where the complex has highest stability (c.a. 12.4 kcal/mol in relative to the reactant complex) as shown in Figure 32. The final product complex is stabilized by the hydrogen bonding interactions with the positive charged methylammonium of Lys65 and many bound water molecules. However, it can be seen that this final product complex has higher energy compared to that of the previous simulation system (Model A) by approximate energy 6 kcal/mol. Only the final product complex shows the carboxylic acid of Asp185 flipping to interact with the Asp186 carboxylate group via a hydrogen bonding interaction. The energies for reverse reactions from the final product to the reactant complexes in Model B are 10.0, 1.5 and 0 kcal/mol (see Figure 27 and 30). In latter, it indicates that the H-transfer reaction from the OD1-oxygen of Asp185 back to the 3'-hydroxyl group of the terminal primer is a spontaneous reaction.

In Asp186 as a base mechanism, the energy barrier for the reactions of nucleophilic attack and Pα-O3α departure are c.a. 13.3 and 5.3 kcal/mol, respectively, as the free energy profile for this reaction pathway presented in Figure 30. During the simulations of the driven reaction from the starting structure to the final product complexes, the creation of pentacoordinated intermediate is also found at 0 Å of the reaction coordinate similar to all previous results in either Model A or Model B.

Whereas the final product complex is located in range of reaction coordinates larger than 2.4 Å. The energies for reverse reactions in Asp186 as a base mechanism from the final product to the reactant complexes in Model B are 9.0, 2.9 and 2.7 kcal/mol (see Figure 27 and 30).

Consequently, in comparison between two different aspartates as a base mechanism, the most feasible reaction mechanism in the mono-protonated system (Model B) is when the Asp185 has its function as a hydrogen acceptor for deprotonation of the 3'-OH group on the primer terminus. Within this reaction pathway, the energy barrier for overall reaction is of 22.4 kcal/mol while the rate-limiting step is the H3T-proton transferring to the Asp185 carboxylate side chain with 15.6 kcal/mol of activation energy barrier.

Figure 30   Free energy profiles for nucleophilic addition in reaction coordinate as: different reaction coordinate between Pα-O3α breaking and Pα-O3' forming bond.

87
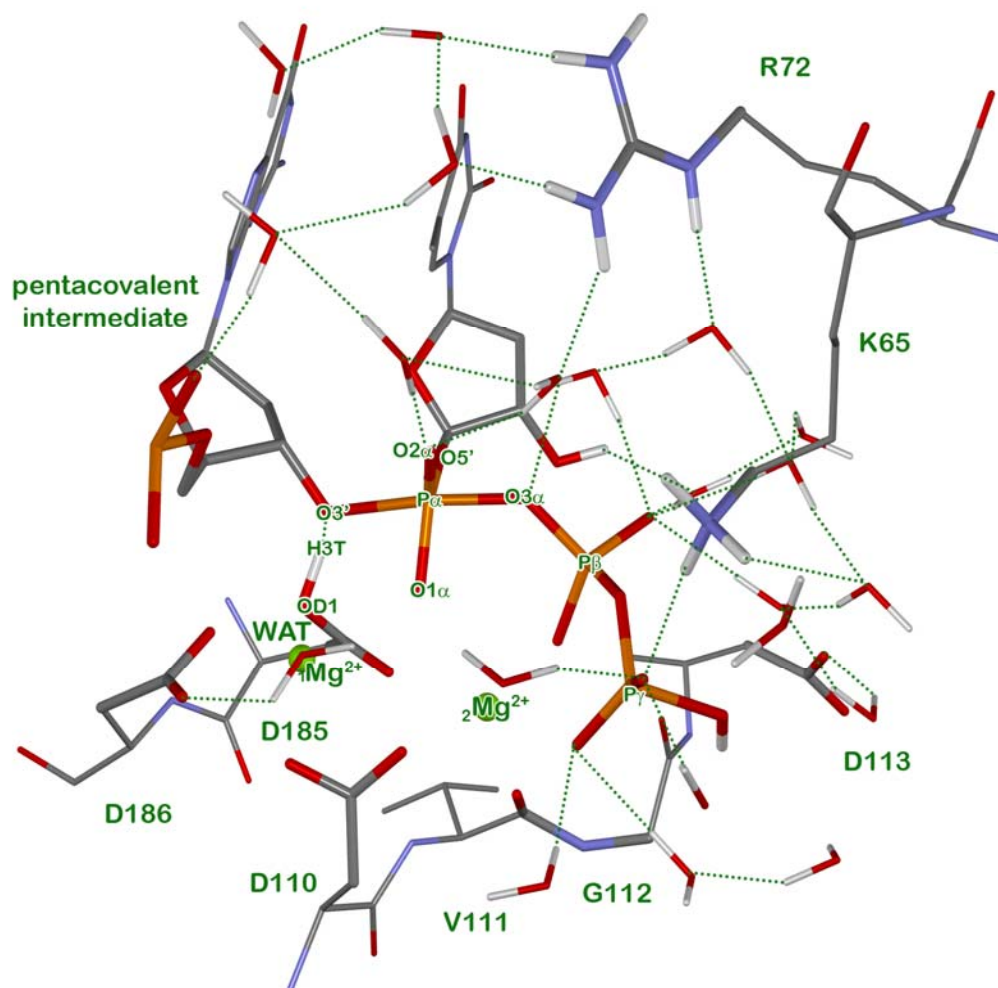
Figure 31   Pentacovalent intermediate (Model B), magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.
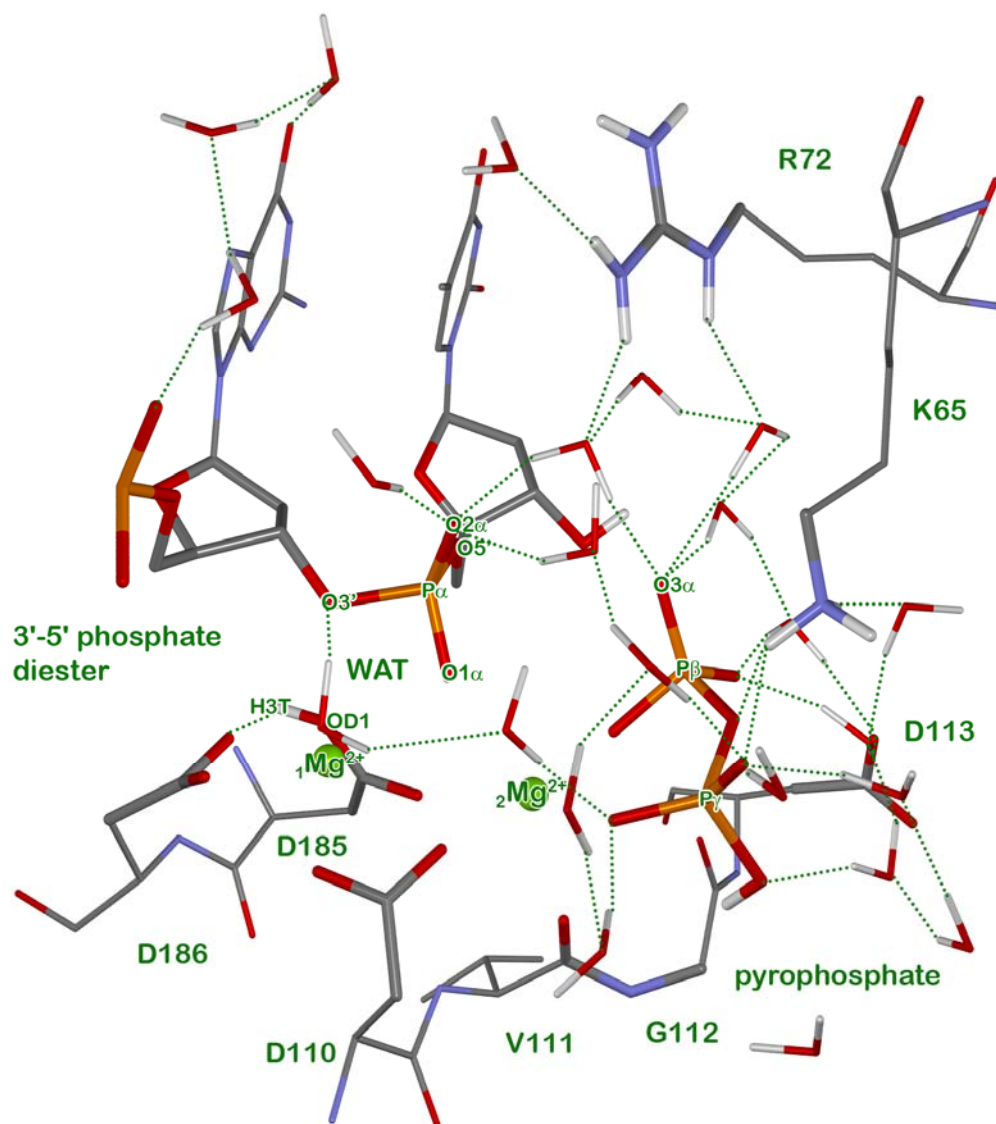
Figure 32 Final product complex, magnesium coordination environment and possible hydrogen bonds (dot line) taken from PM3 QM/MM simulation snapshot. Only H atoms involved the hydrogen bonding interaction and H atoms on water molecules are shown.

From all simulated results in both fully deprotonated and mono-protonated systems (Models A and B), the most favorable reaction mechanism in the HIV-1 RT active site is predicted to involve the H-transfer reaction to Asp185 according to nucleotide addition with overall activation barrier of 18.4 kcal/mol in Model A. It is noteworthy that the fully deprotonated model can give more reasonable results involving the mechanistic reaction and the critical structures along the reaction pathway compared to the mono-protonated model. In addition, Lys65 and Arg72 in HIV-1 RT play important roles in stabilizing the dTTP substrate via hydrogen bond interactions with its triphosphate moiety along the reaction pathway.

CONCLUSION

MM molecular dynamics simulations, and QM/MM optimizations, provide insight into the structure of the catalytically relevant substrate complex of HIV-1 reverse transcriptase. Modelling allows the catalytically relevant complex, to be studied, extending beyond the crystal structure of the catalytic HIV-1 RT/DNA/dTTP ternary complex, in which the terminal 3'-OH of the primer is missing.

We have tested different protonation states of the γ-phosphate of the dTTP substrate (protonation in two different positions) and two active-site aspartates (Asp185 and Asp186). Comparison with the crystal structure indicates that the models with both asparates charged are more probable. In simulations on all models containing protonated aspartic acid residues, the active site became significantly distorted from the original crystal geometry. Our results also indicate that binding with the deprotonated triphosphate of the dTTP substrate is tighter and more stable than with the mono-protonated form. Important hydrogen bonds to the phosphates come from the Lys65, Lys72, Asp113 and Ala114 residues and bound water molecules. Thus, two simulated structures in difference of the triphosphate protonation states with two aspartates in their basic form (Models A and B) are acceptable to be the representation of the polymerase active site serving as an initial structure for the QM/MM minimization.

The simulations support a characteristic polymerase active site, with approximately octahedral coordination environment of the divalent cations. Coordinating the catalytic $_1Mg^{2+}$ as ligands are the 3'-OH terminal primer, α-phosphate, the sidechains of three aspartates (Asp110, Asp185 and Asp186) and a modelled water molecule. The ligands of the other (bridging) magnesium ($_2Mg^{2+}$) are three phosphate oxygens, the sidechains of two bridging aspartates (Asp110 and Asp185) and a backbone carbonyl of valine (Val111).

Two different QM/MM methods (AM1/CHARMM and PM3/CHARMM) were applied to study the structure of the active site. Generally, both methods gave

structures in good agreement with the crystal structure, and with the structures derived from MM modelling. There were some small differences in the key active-site structural parameters (for example, the Mg-O and P-O distances and the internuclear separation between divalent cations) calculated by the two QM/MM techniques. However, they both gave realistic structures of the dTTP triphosphate, in good agreement with the crystal structure. These results suggest that these methods could be usefully applied in future investigations of the chemical mechanism (deprotonation and DNA polymerization) of HIV-1 reverse transcriptase. Analysis of the treatment of the challenging chemistry involved by these semiempirical methods would also clearly be required.

In this study, three different base mechanisms for H-transfer reaction following by nucleotide addition in the HIV-1 RT active site were estimated with two different semiempirical QM/MM methods (AM1/CHARMM and PM3/CHARMM). Three different mechanisms for the first reaction (Step 1) of H-transferring from the 3'-hydroxyl group of the terminal primer, are investigated based on two different systems (the fully deprotonated dTTP: Model A and the mono-protonated dTTP: Model B): (i) proton transfer to the OD1-carboxylate oxygen of the Asp185 residue, (ii) proton transfer to the OD1-carboxylate oxygen of the Asp186 residue, and (iii) proton transfer to the O5'-oxygen on α-phosphate of the dTTP substrate. The most feasible H-transfer reaction was found to proceed in a corresponding reaction path via the Asp185 residue which plays an important role as the catalytic base in both models A and B representing for the HIV-1 RT. Consequently, the nucleophilic attacking on α-phosphorus of the dTTP substrate (Step 2) leading to the formation of the pentacovalent intermediate and the subsequent Pα-O3α breaking bond of the pentacoordinated structure (Step 3). 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 consisting of three mechanistic steps is energetically closed to 18.4 kcal/mol in which the rate-limiting step is the H-transfer reaction to Asp185 residue found in model of the deprotonated dTTP substrate. The inversion character on α-phosphorus atom of dTTP substrate from the reactant to the final

product via a nearly perfect planar configuration of pentacoordinate intermediate indicates the S_N2 reaction of the nucleotide addition. The critical structures along the reaction pathway including the reactant, product, intermediate and final product complexes have been stabilized by the hydrogen bonding interactions with the methylammonium of Lys65, the guanidinium of Arg72 and the backbone (NH) of Asp113 and some tightly bound water molecules (modelled waters). These studies highlight the utility of QM/MM molecular dynamics simulations for calculation of free energy profiles for enzyme reactions. The resulting mechanistic insight may help in inhibitor design. The results reported here will form a good basis for our continuing investigations of this important HIV enzyme.

LITERATURE CITED

Abashkin, Y.G., J.W. Erickson and S.K. Burt. 2001. Quantum Chemical Investigation of Enzymatic Activity in DNA Polymerase b. A Mechanistic Study. **J. Phys. Chem. B.** 105: 287-292.

Alber, F. and P. Carloni. 2000. Ab initio molecular dynamics studies on HIV-1 reverse transcriptase triphosphate binding site: implications for nucleoside-analog drug resistance. **Protein Science** 9: 2535-2546.

Antonczak, S., G. Monard, M.F. Ruiz-Lopez and J.-L. Rivail. 1998. Modeling of Peptide Hydrolysis by Thermolysin. A Semiempirical and QM/MM Study. **J. Am. Chem. Soc.** 120: 8825-8833.

Bartels, C. and M. Karplus. 1997. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. **J. Comp. Chem.** 18: 1450-1462.

Bassan, A., M.R.A. Blomberg and P.E.M. Siegbahn. 2004. A theoretical study of the cis-dihydroxylation mechanism in naphthalene 1,2-dioxygenase. **J. Biol. Inorg. Chem.** 9: 439-452.

Bathelt, C., M., J. Zurek, J. Mulholland Adrian and N. Harvey Jeremy. 2005. Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. **J. Am. Chem. Soc.** 127: 12900-12908.

Bathelt, C.M., L. Ridder, A.J. Mulholland and J.N. Harvey. 2003. Aromatic Hydroxylation by Cytochrome P450: Model Calculations of Mechanism and Substituent Effects. **J. Am. Chem. Soc.** 125: 15004-15005.

Bathelt, C.M., L. Ridder, A.J. Mulholland and J.N. Harvey. 2004. Mechanism and structure-reactivity relationships for aromatic hydroxylation by cytochrome P450. **Org. Biomol. Chem.** 2: 2998-3005.

Beese, L.S. and T.A. Steitz. 1991. Structural basis for the 39-59 exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. **EMBO J.** 10: 25-33.

Beese, L.S., V. Derbyshire and T.A. Steitz. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. **Science** 260: 352-355.

Bibillo, A., D. Lener, G.J. Klarmann and S.F.J. Le Grice. 2005. Functional roles of carboxylate residues comprising the DNA polymerase active site triad of Ty3 reverse transcriptase. **Nucleic Acids Res.** 33: 171-181.

Borowski, T., A. Bassan and P.E.M. Siegbahn. 2004. 4-Hydroxyphenylpyruvate Dioxygenase: A Hybrid Density Functional Study of the Catalytic Reaction Mechanism. **Biochemistry** 43: 12331-12342.

Brody, R.S., S. Adler, P. Modrich, W.J. Stec, P.A. Frey and Z.J. Leznikowski. 1982. Stereochemical course of nucleotidyl transfer catalyzed by bacteriophage T7 induced DNA polymerase. **Biochemistry** 21: 2570-2572.

Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus. 1983a. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. **J. Comp. Chem.** 4: 187-217.

Brooks, C.L. and M. Karplus. 1983b. Deformable stochastic boundaries in molecular dynamics. **J. Chem. Phys.** 79: 6312-6325.

Brooks, C.L. and M. Karplus. 1989. Solvent effects on protein motion and protein effects on solvent motion: Dynamics of the active site of lysozyme. **J. Mol. Biol.** 208: 159-181.

Brunger, A.T. and M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. **Proteins:Struct., Funct., Genet.** 4: 148-156.

Burgers, P.M.J. and F. Eckstein. 1979. A study of the mechanism of DNA polymerase I from Escherichia coli with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. **J. Biol. Chem.** 254: 6889-6893.

Chandrasekaran, A., N.V. Timosheva, R.O. Day and R.R. Holmes. 2003. Influence of Hydrogen Bonding in Competition with Lattice Interactions on Carbonyl Coordination at Phosphorus. Implications for Phosphoryl Transfer Activated States. **Inorg. Chem.** 42: 3285-3292.

Cheng, Y., Y. Zhang and J.A. McCammon. 2005. How Does the cAMP-Dependent Protein Kinase Catalyze the Phosphorylation Reaction: An ab Initio QM/MM Study. **J. Am. Chem. Soc.** 127: 1553-1562.

Cho, K.-B., V. Pelmenschikov, A. Graeslund and P.E.M. Siegbahn. 2004. Density Functional Calculations on Class III Ribonucleotide Reductase: Substrate Reaction Mechanism with Two Formates. **J. Phys. Chem. B.** 108: 2056-2065.

Chong, Y., K. Borroto-Esoda, A. Furman Phillip, F. Schinazi Raymond and K. Chu Chung. 2002. Molecular mechanism of DApd/DXG against zidovudine- and lamivudine- drug resistant mutants: a molecular modelling approach. **Antivir. Chem. Chemother.** 13: 115-128.

Cisneros, G.A., M. Wang, P. Silinski, M.C. Fitzgerald and W. Yang. 2004. The Protein Backbone Makes Important Contributions to 4-Oxalocrotonate Tautomerase Enzyme Catalysis: Understanding from Theory and Experiment. **Biochemistry** 43: 6885-6892.

Cummins, P.L. and J.E. Gready. 2000. Combined quantum and molecular mechanics (QM/MM) study of the ionization state of 8–methylpterin substrate bound to dihydrofolate reductase. **J. Phys. Chem. B.** 104: 4503–4510.

Derbyshire, V., N.D.F. Grindley and C.M. Joyce. 1991. The 39-59 exonuclease of DNA polymerase I of Escherichia coli: contribution of each amino acid at the active site to the reaction. **EMBO J.** 10: 17-24.

Dewar, M.J.S., E.G. Zoebisch, E.F. Healy and J.J.P. Stewart. 1985. AM1: a new general purpose quantum mechanical molecular model. **J. Am.Chem. Soc.** 107: 3902-3909.

Dewar, M.J.S. and C. Jie. 1989. AM1 parameters for phosphorus. **THEOCHEM-J. Mol. Struct.** 56: 1-13.

Dinner, A.R., G.M. Blackburn and M. Karplus. 2001. Uracil-DNA glycosylase acts by substrate autocatalysis. **Nature (London, United Kingdom)** 413: 752-755.

Doublie, S., S. Tabor, A.M. Long, C.C. Richardson and T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. **Nature** 391: 251-258.

Feierberg, I., V. Luzhkov and J. Aqvist. 2000. Computer simulation of primary kinetic isotope effects in the proposed rate-limiting step of the glyoxalase I catalyzed reaction. **J. Biol. Chem.** 275: 22657-22662.

Field, M.J., P.A. Bash and M. Karplus. 1990. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. **J. Comp. Chem.** 11: 700-733.

Florian, J., M.F. Goodman and A. Warshel. 2003. Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. **J. Am. Chem. Soc.** 125: 8163-8177.

Foloppe, N. and A.D. Mackerell. 2000. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. **J. Comp. Chem.** 21: 86-104.

Fuxreiter, M. and A. Warshel. 1998. Origin of the Catalytic Power of Acetylcholinesterase: Computer Simulation Studies. **J. Am. Chem. Soc.** 120: 183–194.

Gao, H.Q., P.L. Boyer, S.G. Sarafianos, E. Arnold and S.H. Hughes. 2000. The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase. **J. Mol. Biol.** 300: 403-418.

Gleeson, M.P., H. Hillier Ian and A. Burton Neil. 2004. Theoretical analysis of peptidyl alpha-ketoheterocyclic inhibitors of human neutrophil elastase: Insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. **Org. Biomol. Chem.** 2: 2275-2280.

Glennon, T.M. and A. Warshel. 1998. Energetics of the Catalytic Reaction of Ribonuclease A: A Computational Study of Alternative Mechanisms. **J. Am. Chem. Soc.** 120: 10234-10247.

Glennon, T.M., J. Villa and A. Warshel. 2000. How Does GAP Catalyze the GTPase Reaction of Ras?: A Computer Simulation Study. **Biochemistry** 39: 9641-9651.

Gregersen, B.A., X. Lopez and D.M. York. 2003. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions. **J. Am. Chem. Soc.** 125: 7178-7179.

Hannongbua, S., K. Nivesanond, L. Lawtrakul, P. Pungpo and P. Wolschann. 2001. 3D-Quantitative Structure-Activity Relationships of HEPT Derivatives as HIV-1 Reverse Transcriptase Inhibitors, Based on Ab Initio Calculations. **J. Chem. Inf. Model.** 41: 848-855.

Harris, D., N. Kaushik, P.K. Pandey, P.N. Yadav and V.N. Pandey. 1998. Functional analysis of amino acid residues constituting the dNTP binding pocket of HIV-1 reverse transcriptase. **J. Biol. Chem.** 273: 33624-33634.

Hermann, J.C., C. Hensen, L. Ridder, A.J. Mulholland and H.-D. Hoeltje. 2005. Mechanisms of Antibiotic Resistance: QM/MM Modeling of the Acylation Reaction of a Class A b-Lactamase with Benzylpenicillin. **J. Am. Chem. Soc.** 127: 4454-4465.

Hostomsky, Z., Z. Hostomska, T.B. Fu and J. Taylor. 1992. Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. **J. Virol.** 66: 3179-3182.

Huang, H., R. Chopra, G.L. Verdine and S.C. Harrison. 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. **Science** 282: 1669-1675.

Hutter, M.C., J.R. Reimers and N.S. Hush. 1998. Modeling the bacterial photosynthetic reaction center. 1. Magnesium parameters for the semiempirical AM1 method developed using a genetic algorithm. **J. Phys. Chem. B.** 102: 8080-8090.

Jacobo-Molina, A., J. Ding, R.G. Nanni, A.D. Clark, Jr., X. Lu, C. Tantillo, R.L. Williams, G. Kamer, A.L. Ferris and P. Clark. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. **Proc. Natl. Acad. Sci. U.S.A.** 90: 6320-6324.

Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. **J. Chem. Phys.** 79: 926-935.

Kaushik, N., V.N. Pandey and M.J. Modak. 1996. Significance of the O-helix residues of Escherichia coli DNA polymerase I in DNA synthesis: dynamics of the dNTP binding pocket. **Biochemistry** 35: 7256-7266.

Kedar, P.S., J. Abbotts, T. Kovacs, K. Lesiak, P. Torrence and S.H. Wilson. 1990. Mechanism of HIV reverse transcriptase: enzyme-primer interaction as revealed through studies of a dNTP analogue, 3'-azido-dTTP. **Biochemistry** 29: 3603-3611.

Khandogin, J. and D.M. York. 2004. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. **Proteins** 56: 724-737.

Kharyin, N.A., J.P. Synder and F.M. Menger. 1999. Mechanism of chorismate mutase: Contribution of conformational restriction to catalysis in the claisen rearrangement. **J. Am. Chem. Soc.** 121: 11831–11846.

Klarmann, G.J., R.A. Smith, R.F. Schinazi, T.W. North and B.D. Preston. 2000. Site-specific incorporation of nucleoside analogs by HIV-1 reverse transcriptase and the template grip mutant P157S. Template interactions influence substrate recognition at the polymerase active site. **J. Biol. Chem.** 275: 359-366.

Kollman, P.A., B. Kuhn, O. Donini, M. Perakyla, R. Stanton and D. Bakowies. 2001. Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: quantum mechanical-free energy calculations on chemical reactions in enzymes and in aqueous solution. **Acc. Chem. Res.** 34: 72-79.

Kolmodin, K. and J. Aqvist. 1999a. Computational modeling of the rate limiting step in low molecular weight protein tyrosine phosphatase. **FEBS Letters** 456: 301-305.

Kolmodin, K., P. Nordlund and J. Aqvist. 1999b. Mechanism of substrate dephosphorylation in low Mr protein tyrosine phosphatase. **Proteins:Struct., Funct., Genet.** 36: 370-379.

Kuhn, B. and P.A. Kollman. 2000. QM–FE and molecular dynamics calculations on catechol o–methyltransferase: Free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme–catalyzed reaction. **J. Am. Chem. Soc.** 122: 2586–2596.

Kumar, S., D. Bouzida, R.H. Swendsen, P.A. Kollman and J.M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. **J. Comp. Chem.** 13: 1011-1021.

Lau, E.Y. and T.C. Bruice. 1998. Importance of correlated motions in forming highly reactive near attack conformations in catechol o-methyltransferase. **J. Am. Chem. Soc.** 120: 12387–12394.

Lau, E.Y. and T.C. Bruice. 1999. Active site dynamics of the HhaI methyltransferase: insights from computer simulation. **J. Mol. Biol.** 293: 9-18.

Lau, E.Y., K. Kahn, P.A. Bash and T.C. Bruice. 2000. The importance of reactant positioning in enzyme catalysis: a hybrid quantum mechanics/molecular mechanics study of a haloalkane dehalogenase. **Proceedings of the National Academy of Sciences of the United States of America** 97: 9937-9942.

Lawtrakul, L., A. Beyer, S. Hannongbua and P. Wolschann. 2004. Quantitative Structural Rearrangement of HIV-1 Reverse Transcriptase on Binding to Non-Nucleoside Inhibitors. **Monatsh. Chemie** 135: 1033-1046.

Li, G. and Q. Cui. 2003. What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. **J. Am. Chem. Soc.** 125: 15028–15038.

Liu, H., Y. Zhang and W. Yang. 2000. How is the active site of enolase organized to catalyze two different reaction steps? **J. Am. Chem. Soc.** 122: 6560-6570.

Lopez, X. and D.M. York. 2003. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. **Theoretical Chemistry Accounts** 109: 149-159.

Lu, D. and G. Voth. 1998. Proton Transfer in the Enzyme Carbonic Anhydrase: An ab initio Study. **J. Am. Chem. Soc.** 120: 4006–4014.

Lyne, P.D., M. Hodoscek and M. Karplus. 1999. A Hybrid QM-MM Potential Employing Hartree-Fock or Density Functional Methods in the Quantum Region. **J. Phys. Chem. A.** 103: 3462-3471.

MacKerell, A.D., J. Wiorkiewicz-Kuczera and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. **J. Am. Chem. Soc.** 117: 11946-11975.

MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. **J. Phys. Chem. B.** 102: 3586-3616.

Mackerell, A.D. and N.K. Banavali. 2000. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. **J. Comp. Chem.** 21: 105-120.

Marti, S., J. Andres, V. Moliner, E. Silla, I. Tunon, J. Bertran and M.J. Field. 2001. A hybrid potential reaction path and free energy study of the chorismate mutase reaction. **J. Am. Chem. Soc.** 123: 1709–1712.

Merz Jr., K.M. and L. Banci. 1997. Binding of bicarbonate to human carbonic anhydrase II: A continuum of binding states. **J. Am. Chem. Soc.** 119: 863–871.

Meyer, P.R., S.E. Matsuura, A.G. So and W.A. Scott. 1998. Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. **Proc. Natl. Acad. Sci. U. S. A.** 95: 13471-13476.

Morikis, D., A.H. Elcock, P.A. Jennings and J.A. McCammon. 2001. Proton transfer dynamics of GART: the pH-dependent catalytic mechanism examined by electrostatic calculations. **Protein Sci.** 10: 2379-2392.

Morikis, D., A.H. Elcock, P.A. Jennings and J.A. McCammon. 2003. The pH dependence of stability of the activation helix and the catalytic site of GART. **Biophys. Chem.** 105: 279-291.

Mulholland, A.J. and W.G. Richards. 1998. Modeling Enzyme Reaction Intermediates and Transition States: Citrate Synthase. **J. Phys. Chem. B.** 102: 6635-6646.

Mulholland, A.J., P.D. Lyne and M. Karplus. 2000. Ab Initio QM/MM Study of the Citrate Synthase Mechanism. A Low-Barrier Hydrogen Bond Is not Involved. **J. Am. Chem. Soc.** 122: 534-535.

Mulholland, A.J. 2005. Modelling enzyme reaction mechanisms, specificity and catalysis. **Drug discovery today** 10: 1393-1402.

Olsson, M.H.M., P.K. Sharma and A. Warshel. 2005. Simulating redox coupled proton transfer in cytochrome c oxidase: Looking for the proton bottleneck. **FEBS Lett.** 579: 2026–2034.

Park, H., N. Brothers Edward and M. Merz Kenneth, Jr. 2005. Hybrid QM/MM and DFT investigations of the catalytic mechanism and inhibition of the dinuclear zinc metallo-beta-lactamase CcrA from Bacteroides fragilis. **J. Am. Chem. Soc.** 127: 4232-4241.

Ranaghan, K.E., L. Ridder, B. Szefczyk, W.A. Sokalski, J.C. Hermann and A.J. Mulholland. 2003. Insights into enzyme catalysis from QM/MM modelling: transition state stabilization in chorismate mutase. **Mol. Phys.** 101: 2695-2714.

Ranaghan, K.E., L. Ridder, B. Szefczyk, W.A. Sokalski, J.C. Hermann and A.J. Mulholland. 2004. Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. **Org. Biomol. Chem.** 2: 968-980.

Reuter, N., A. Dejaegere, B. Maigret and M. Karplus. 2000. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. **J. Phys. Chem. A.** 104: 1720-1735.

Ridder, L., A.J. Mulholland, I.M.C.M. Rietjens and J. Vervoort. 2000. A Quantum Mechanical/Molecular Mechanical Study of the Hydroxylation of Phenol and Halogenated Derivatives by Phenol Hydroxylase. **J. Am. Chem. Soc.** 122: 8728-8738.

Ridder, L., I.M.C.M. Rietjens, J. Vervoort and A.J. Mulholland. 2002. Quantum mechanical/molecular mechanical free energy simulations of the glutathione S-transferase (M1-1) reaction with phenanthrene 9,10-oxide. **J. Am. Chem. Soc.** 124: 9926-9936.

Ridder, L., J.N. Harvey, I.M.C.M. Rietjens, J. Vervoort and A.J. Mulholland. 2003a. Ab Initio QM/MM Modeling of the Hydroxylation Step in p-Hydroxybenzoate Hydroxylase. **J. Phys. Chem. B.** 107: 2118-2126.

Ridder, L. and J. Mulholland Adrian. 2003b. Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity. **Curr. Top. Med. Chem.** 3: 1241-1256.

Rungrotmongkol, T., S. Hannongbua and A. Mulholland. 2004. Mechanistic study of HIV-1 reverse transcriptase at the active site based on QM/MM method. **J. Theor. Comput. Chem.** 3: 491-500.

Ryckaert, J.P., G. Ciccotti and H.J.C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **J. Comput. Phys.** 23: 327-341.

Saenger, W. 1983. Principles of nucleic acid structure. **NewYork: Springer-Verlag**

Saen-Oon, S., S. Hannongbua and P. Wolschann. 2003. Structural flexibility of non-nucleoside HIV-1 reverse transcriptase inhibitor: 9-Cl TIBO as explained by potential energy surface and (13)C and (1)H NMR calculations, based on ab initio and density functional study. **J. Chem. Inf. Model.** 43: 1412-1422.

Sarafianos, S.G., V.N. Pandey, N. Kaushik and M.J. Modak. 1995. Glutamine 151 participates in the substrate dNTP binding function of HIV-1 reverse transcriptase. **Biochemistry** 34: 7207-7216.

Sawaya, M.R., R. Prasad, S.H. Wilson, J. Kraut and H. Pelletier. 1997. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. **Biochemistry** 36: 11205-11215.

Schwarzl, S.M., J.C. Smith and S. Fischer. 2006. Insights into the Chemomechanical Coupling of the Myosin Motor from Simulation of Its ATP Hydrolysis Mechanism. **Biochemistry** 45: 5830-5847.

Sheppard, D.W., N.A. Burton and I.H. Hillier. 2000. Ab initio hybrid quantum mechanical/molecular mechanical studies of the mechanisms of the enzymes protein kinase and thymidine phosphorylase. **THEOCHEM-J. Mol. Struct.** 506: 35-44.

Sluis-Cremer, N., D. Arion, N. Kaushik, H. Lim and M.A. Parniak. 2000. Mutational analysis of Lys65 of HIV-1 reverse transcriptase. **Biochem. J.** 348: 77-82.

Stanton, R.V., D.S. Hartsough and J. K.M. Merz. 1995. An examination of a density functional/molecular mechanical coupled potential. **J. Comput. Chem.** 16: 113-128.

Steitz, T.A. 1999. DNA polymerases: structural diversity and common mechanisms. **J. Biol. Chem.** 274: 17395-17398.

Stewart, J.J.P. 1989. Optimization of parameters for semiempirical methods. I. Method. **J. Comput. Chem.** 10: 209-220.

Stewart, J.J.P. 1991. Optimization of parameters for semiempirical methods. III. Extension of PM3 to beryllium, magnesium, zinc, gallium, germanium, arsenic, selenium, cadmium, indium, tin, antimony, tellurium, mercury, thallium, lead, and bismuth. **J. Comput. Chem.** 12: 320-341.

Swamy, K.C.K. and N.S. Kumar. 2003. Pentacoordinated phosphorus in action. **Curr. Sci.** 85: 1256–1258.

Toba, S., G. Colombo and K.M. Merz Jr. 2000. Solvent Dynamics and Mechanism of Proton Transfer in Human Carbonic Anhydrase II. **J. Am. Chem. Soc.** 121: 2290–2302.

Torres, R.A., B. Schiott and T.C. Bruice. 1999. Molecular dynamics simulations of ground and transition states for the hydride transfer from formate to NAD+ in the active site of formate dehydrogenase. **J. Am. Chem. Soc.** 121: 8164-8173.

Van Der Vaart, A., V. Gogonea, S.L. Dixon and K.M. Merz, Jr. 2000. Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method. **J. Comp. Chem.** 21: 1494-1504.

Van Gunsteren, W.F. and H.J.C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. **Mol. Phys.** 34: 1311-1327.

Varnai, P. and A. Warshel. 2000. Computer Simulation Studies of the Catalytic Mechanism of Human Aldose Reductase. **J. Am. Chem. Soc.** 122: 3849–3860.

Wang, J., A.K. Sattar, C.C. Wang, J.D. Karam, W.H. Konigsberg and T.A. Steitz. 1997. Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. **Cell** 89: 1087-1099.

Warshel, A. and M. Levitt. 1976. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. **J. Mol.Biol.** 103: 227-249.

Warshel, A., M. Strajbl, J. Villa and J. Florian. 2000. Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization. **Biochemistry** 39: 14728-14738.

Wondrak, E.M., J. Loewer and R. Kurth. 1986. Functional purification and enzymic characterization of the RNA dependent DNA polymerase of human immunodeficiency virus. **J. Gen. Virol.** 67: 2791-2797.

Xu, D., Y. Wei, J. Wu, D. Dunaway-Mariano, H. Guo, Q. Cui and J. Gao. 2004. QM/MM Studies of the Enzyme-Catalyzed Dechlorination of 4-Chlorobenzoyl-CoA Provide Insight into Reaction Energetics. **J. Am. Chem. Soc.** 126: 13649-13658.

Yang, L., K. Arora, W.A. Beard, S.H. Wilson and T. Schlick. 2004. Critical Role of Magnesium Ions in DNA Polymerase b's Closing and Active Site Assembly. **J. Am. Chem. Soc.** 126: 8441-8453.

Zheng, Y.J. and T.C. Bruice. 1997. A theoretical examination of the factors controlling the catalytic efficiency of a transmethylation enzyme: Catechol o–methyltransferase. **J. Am. Chem. Soc.** 119: 8137–8145.

Zurek, J., A.L. Bowman, W.A. Sokalski and A.J. Mulholland. 2004. MM and QM/MM Modeling of Threonyl-tRNA Synthetase: Model Testing and Simulations. **Struct. Chem.** 15: 405-414.

APPENDIX

APPENDIX A: Theoretical Background

<u>Theoretical Background in Quantum Chemistry</u>

<u>1. Quantum mechanics</u>

Molecular orbital calculation is the important method in quantum chemistry for approximate structures and dynamics of molecular system. This approach provides a great promise in calculating electronic structures and predicting properties of drug molecules. Up to date, molecular orbital investigations have been introduced into drug research to study mechanisms of action and to guide the design of more potent agents.

The quantum chemical methods are based on finding solutions to the time independent Schrödinger wave equation on molecular orbital theory. In its simplest form, the time-independent Schrödinger equation is expressed as

$$H\Psi \quad = \quad E\Psi \tag{1}$$

The rewritten time-independent Schrödinger equation in equation (1) can be shown as:

$$H(r, R)\Psi(r, R) \quad = \quad E(R)\Psi(r, R) \tag{2}$$

The molecular wavefunction, $\Psi(r, R)$, depends on explicitly on the coordinates of all electrons in the system (r) and implicitly on the collective nuclear coordinates (R). When the wave function (Ψ) is known, all physical observables of interest can be calculated. Mostly, the time-independent Schrödinger equation becomes too complicated to be solved directly. Consequently, to make calculation of molecular properties possible, approximations were introduced.

1.1 Born-Oppenheimer Approximation

The Born-Oppenheimer approximation allows for the implicit on the collective nuclear coordinates (R) rather than the explicit on the electron coordinates (r) of the system. Consequently, the molecular wavefunction extremely depends on the nuclei positions as shown in equation (3).

$$\Psi(\text{nuclei, electrons}) \quad = \quad \Psi(\text{electrons})\Psi(\text{nuclei}) \tag{3}$$

Therefore, solving the Schrödinger equation yields the energy of the system which depends on the nuclei positions, E(R). H is the Hamiltonian operator which gives the kinetic and potential energies of the system,

$$H = T + V \tag{4}$$

Then, rewrite equation (1) is;

$$\left\{ \frac{-\bar{h}^2}{2m_e} \nabla^2 + V \right\} \Psi = E\Psi \tag{5}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{6}$$

ħ is Plank's constant divided by 2π. Ψ is denoted for the wavefunction which characterizes the particle's properties. E is the energy of the particle.

The first term of the left-hand side of the equation (3) represents the kinetic energy (T) of each electron, of mass m_e, moving through space in the x, y and z directions.

The second term of Hamiltonian operator is potential energy (V) which is comprised the electron-nuclear attractions, electron-electron repulsions and nuclear-nuclear repulsions within the system. Therefore, the total energy of the system can be written as:

$$E \ = \ \frac{\int \psi^* H \, \psi \ d\tau}{\int \psi^* \, \psi \ d\tau} \tag{7}$$

where ψ^* is the complex conjugate of the wavefunction and $d\tau$ is the integration over all space.

1.2 Orbital Approximation

The exact solution of the Schrödinger equation is only possible for one electron systems. Only the Born-Oppenheimer approximation as mentioned above can not solve the exact solution of the molecular wave function and energy of a system. The manyelectron wave-function Ψ is approximated as a product of one-electron functions as called the orbital approximation.

$$\Psi_{electronic} \quad = \quad | \ \psi_1 \ \psi_2 \ \psi_3 \ \dots \ \psi_N \ | \tag{8}$$

$$= \quad O(s) \ A \ [\psi_1 \ \psi_2 \ \psi_3 \ \dots \ \psi_N] \tag{9}$$

where A is the antisymmetrizer, ensuring the wave function obeys the Pauli exclusion principle, and O(s) is a spin projection operator that ensures that the wave function remains an eigenfunction of the spin-squared operator, S^2. It has the advantage of satisfying the condition that the overall wavefuntion must be antisymmetric under electron exchange. It indicates that the wavefunction must change if the space and spin coordinates of two electrons are interchanged.

1.3 The LCAO-MO Approximation

For a molecular system, the approximate molecular orbital Ψ_i is customarily expanded as a linear combination of atomic orbitals, ϕ, known as the Linear Combination of Atomic Orbitals (LCAO) approximation.

$$\Psi_i = \sum_{\mu}^{L} c_{\mu i} \phi_{\mu} \tag{10}$$

where $C_{\mu i}$ are the weighting coefficients and ϕ_{μ} are real atomic functions. The atomic orbitals are normally known as basis functions. The molecular orbital is made up of L basis functions in what is known as basis set.

The requirement that the orbitals are orthonormal is

$$\sum_{\mu\nu} c_{\mu i}^{*} c_{\nu i} S_{\mu\nu} = \delta_{ij} \tag{11}$$

where δ_{ij} is the Kronecker delta and S_{μ} is overlap integral for atomic functions ϕ_{μ} and ϕ_{ν}

$$S_{\mu\nu} = \int \phi_{\mu}(1) \phi_{\nu}(1) d\tau \tag{12}$$

1.4 Solving for the Molecular Orbital: LCAO-MO- SCF

Introducing Eq. (8) and (9) into Eq. (1), the equation takes the final form generally known as the Roothaan equations as

$$\sum_{\nu} \left(F_{\mu\nu} - \varepsilon_i S_{\mu\nu} \right) c_{\nu i} = 0 \tag{13}$$

The elements of the matrix representation of the Hartree-Fock Hamiltonian operator F are

$$F_{\mu\nu} - H_{\mu\nu}^{core} + \sum P_{\lambda\sigma} \left[(\mu\nu / \lambda\sigma) - \tfrac{1}{2}(\mu\lambda / \nu\sigma) \right] \tag{14}$$

and density matrix defined as

$$P_{\mu\nu} = 2 \sum_i c_{\mu i}^* c_{\nu i} \tag{15}$$

$$(\mu\nu / \lambda\sigma) = \iint \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2) d\tau_1 d\tau_2 \tag{16}$$

and one-electron orbital energy is

$$\varepsilon_i = H_i^{(1)} + \sum_j 2J_{ij} - K_{ij} \tag{17}$$

where

$$\text{Coulomb integral, } J_{ij} = \sum_{\mu\nu\lambda\sigma} c_{\mu i}^* c_{\lambda j}^* c_{\nu i} c_{\sigma j} (\mu\nu / \lambda\sigma) \tag{18}$$

$$\text{and Exchange integral, } K_{ij} = \sum_{\mu\nu\lambda\sigma} c_{\mu i}^* c_{\lambda i}^* c_{\nu i} c_{\sigma i} (\mu\lambda / \nu\sigma) \tag{19}$$

The total electronic energy (ε) is

$$\varepsilon = \sum_i \left(\varepsilon_i + H_{ii}^{(1)} \right) \tag{20}$$

Therefore, Eq. (20) can be written in matrix form as

$$\mathbf{FC = SCE} \tag{21}$$

where E is the diagonal matrix of the ε_i and the elements of a matrix C are the coefficients in the expansion LCAO.

Hartree-Fock or self-consistent field method introduces some elegant approximations to solve a one electron eigenvalue problem, and must be solved iteratively. Solving the equation (19) for the coefficient C describing the LCAO expansion of the orbital ψ_i and orbital energies ε_i which require a matrix diagonalized. Note that F depends on the coefficient C.

They may be usefully transformed by defining new matrices

$$F^\tau = S^{\frac{1}{2}}FS^{-\frac{1}{2}} \tag{22}$$

$$C^\tau = S^{\frac{1}{2}}C \tag{23}$$

then obtain

$$F^\tau C^\tau = C^\tau E \tag{24}$$

Matrix equation (22) can be solved using standard methods. The basis function coefficients can be obtained from C^τ in equation (21). The matrix elements of the Hartree-Fock Hamiltonian operator are dependent on the orbitals through the elements $P_{\mu v}$ and the Roothaan equations are solved by first assuming an initial set of linear expansion coefficients. The whole process is then repeated until the coefficients no longer change within a given tolerance on repeated iteration. The solution is then said to be self-consistent and the method is then referred to as the SCF method. The mathematical steps required for the solution of the Rothaan-Hall equations are outlined in Figure A1.
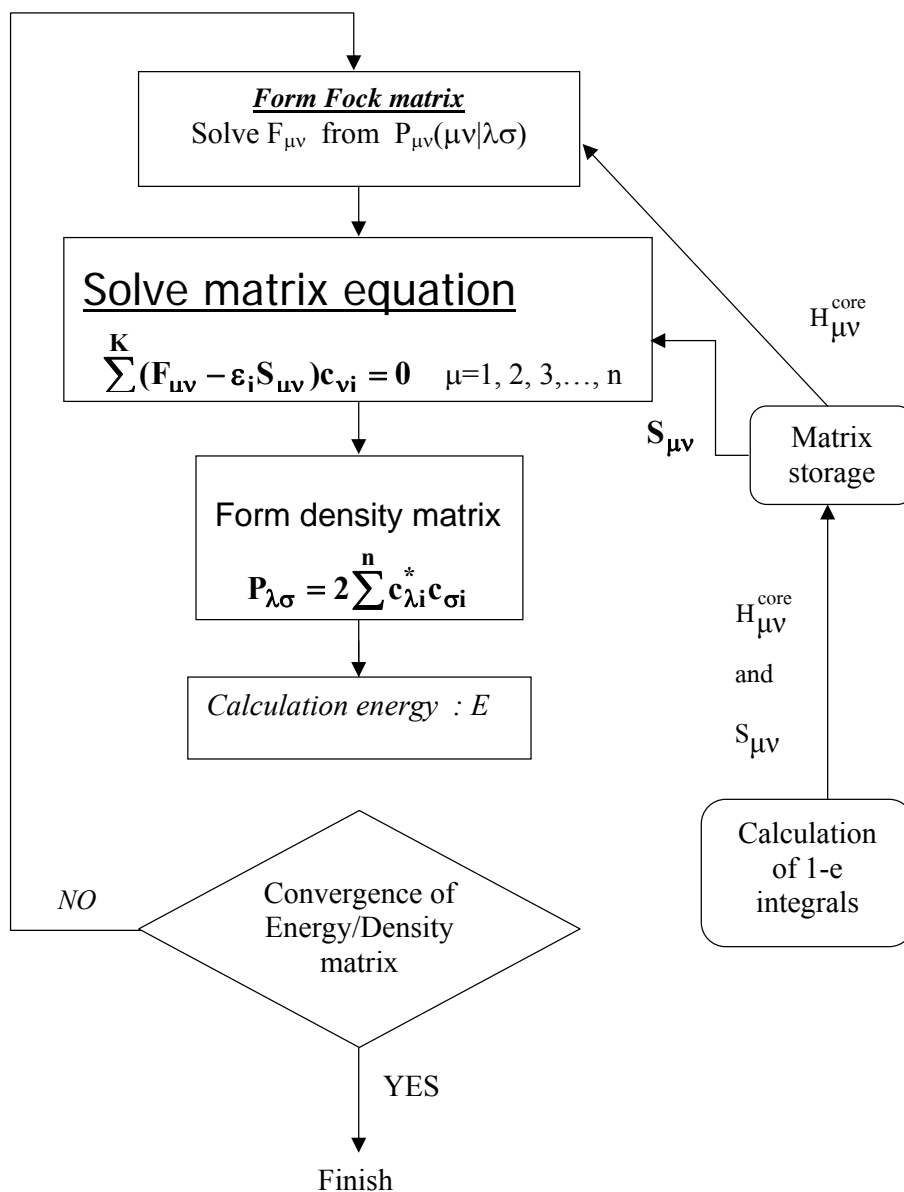
Appendix Figure A1 Sequence of program step required to solve the Roothaan-Hall

equations, Self Consistent Field procedure.

1.5 Semiempirical Methods

There are a large number of semiempirical methods in use today such as MINDO/3, MNDO, AM1, PM3, PM5, and MNDO-*d* which have roughly the same structure. A superficial understanding of these methods and their relationship to *ab initio* methods are important particularly for interpreting the results.

The semiempirical methods have many features in common. They are all self-consistent field (SCF) methods, they take into account electrostatic repulsion and exchange stabilization, and, all calculated integrals are evaluated by approximate means. Further, they all use a restricted basis set of one *s* orbital and three *p* orbitals (p_x, p_y, and p_z) per atom (except MNDO-*d*, which has five *d* orbitals in addition to the *s-p*basis set) and ignore overlap integrals in the secular equation. Thus, instead of solving

$$|H - ES| = 0 \qquad\qquad (25)$$

which *H* is the secular determinant, *S* is the overlap matrix, and *E* is the set of eigenvalues, is solved. These approximations considerably simplify quantum mechanical calculations on systems of chemical interest. As a result, larger systems can be studied. Computational methods are only models, and there is no advantage in rigorously solving Schrödinger's equation for a large system if that system has had to be abbreviated in order to make the calculations tractable. Semiempirical methods are thus seen to be well balanced: they are accurate enough to have useful predictive powers, yet fast enough to allow large systems to be studied.

All the semiempirical methods contain sets of parameters. For MINDO/3 and PM5 atomic and diatomic parameters exist, while MNDO, AM1, PM3, and MNDO-*d* use only single-atom parameters. Not all parameters are optimized for all methods; for example, in MINDO/3, MNDO and AM1 the two electron one center integrals are normally taken from atomic spectra. In the list given in, parameters optimized for a given method are indicated by '*'. A '+' indicates that the value of the parameter was

obtained from experiment (not optimized). Where neither symbol is given, the associated parameter is not used in that method.

All five semiempirical methods also use two experimentally determined constants per atom: the atomic mass of the most abundant isotope and the heat of atomization.

The greatest proportion of the time required to perform an *ab initio* Hartree-Fock SCF calculation is invariably calculating and manipulating integrals. The most obvious way to reduce the computational effort is therefore to neglect or approximate some of these integrals. Semiempirical methods achieve this in part by explicit considering only the valence electrons of the system; the core electrons are assumed into the nuclear core. The overlap matrix, S, is set equal to the identity matrix, I. The main implication of this is that the Roothaan-Hall equations are simplified: FC = SCE becomes FC = CE.

1.5.1 Modified Neglect of Diatomic Overlap (MNDO)

Dewar and Thiel introduced the modified neglect of diatomic overlap (MNDO) method which was based on the neglect of diatomic differential overlap (NDDO); this theory only neglects differential overlap between atomic orbitals on different atoms. The Fock matrix elements in MNDO were as follows:

$$F_{\mu\mu} = H_{\mu\mu}^{core} + \sum_{v\, on\, A}\left[P_{vv}(\mu\mu\,/\,vv) - \frac{1}{2}P_{vv}(\mu v\,/\,\mu v)\right] + \sum_{B \neq A}\sum_{\lambda\, on\, B}\sum_{\sigma\, on\, B} P_{\lambda\sigma}(\mu\mu\,/\,\lambda\sigma) \tag{26}$$

where
$$H_{\mu\mu}^{core} = U_{\mu\mu} - \sum_{B \neq A} V_{\mu\mu B} \tag{27}$$

$$F_{\mu v} = H_{\mu v}^{core} + \frac{3}{2}P_{\mu v}(\mu v\,/\,\mu v) - \frac{1}{2}P_{\mu v}(\mu\mu\,/\,vv) + \sum\sum\sum P_{\lambda\sigma}(\mu v\,/\,\lambda\sigma) \tag{28}$$

$$\mu \text{ and } v \text{ both on A}$$

where
$$H_{\mu\mu}^{core} = -\sum_{B \neq A} V_{\mu\mu B} \qquad (29)$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \frac{3}{2} P_{\mu\nu} (\mu\nu / \mu\nu) - \frac{1}{2} \sum_{\lambda onB} \sum_{\sigma onA} P_{\lambda\sigma} (\mu\sigma / \nu\lambda) \qquad (30)$$

$$\mu \text{ on A and } \nu \text{ on B}$$

where
$$H_{\mu\nu}^{core} = \frac{1}{2} S_{\mu\nu} (\beta_\mu + \beta_\nu) \qquad (31)$$

$V_{\mu\mu B}$ and $V_{\mu\nu B}$ are two-center, one-electron attractions between an electron distribution $\phi_\mu \ \phi_\mu$ or $\phi_\mu \ \phi_\nu$, respectively, on atom A and the core of atom B. These are expressed as follows.

$$V_{\mu\mu B} = Z_B (\mu_A \mu_A / S_A S_A) \qquad (32)$$

$$V_{\mu\nu B} = Z_B (\mu_A \nu_A / S_A S_A) \qquad (33)$$

$$E_{AB} = Z_A Z_B (S_A S_A / S_B S_B)\{1 + \exp(-\alpha_A R_{AB}) + \exp(-\alpha_B R_{AB})\} \qquad (34)$$

1.5.2 Austin Model 1 (AM1) Parametric Method Number 3 (PM3)

The Austin Model 1 (AM1) and PM3 are based on MNDO (the name derives from the fact that PM3 is the third parametrization of MNDO, AM1 being considered the second.) AM1 and PM3 modified the core-core repulsions just outside bonding distances. With this modification the expression for the core-core term was related to the MNDO expression by

$$E_{AB} = E_{MNDO} + \frac{Z_A Z_B}{R_{AB}} \left\{ \begin{array}{l} \sum_I K_{Ai} \exp\left[-L_{Ai}(R_{AB} - M_{Ai})^2\right] \\ + \sum_j K_{Bj} \exp\left[-L_{Bj}(R_{AB} - M_{Bj})^2\right] \end{array} \right\} \tag{35}$$

The extra terms define spherical Gaussian functions, the L, M and K parameters were optimized for each atom. PM3 has two Gaussians per atom, while AM1 has two - four Gaussians per atom. In AM1 a sum of Gaussians is employed to better represent the core repulsion behavior at van der Waals distances.

One advantage of methods parameterised using experimental data is their implicit inclusion of exchange, correlation and vibrational effects. However, it must be kept in mind that these methods are parameterized for molecules in the ground state at 298 K. Therefore, the calculated results are dependence on experimental data means that semiempirical methods would not be expected to perform well on unusual types of molecules for which no data are available from which to construct parameters and should not deal well with systems far from their ground state equilibrium. In addition, the results from semiempirical methods should further be compared to the results of higher-level ab inito or density functional theory.

Concerning to the system of interest (the RT polymerase active site) in this thesis, the QM region contains the deoxythymidine triphosphate (dTTP) which is one of the hypervalent compounds. Therefore, it is important to test the accuracy of semiempirical methods such as AM1 and PM3 when modelling a chemical reaction. They are known to suffer from serious errors in many cases. Relevant for RT is that AM1 treats many phosphorus-containing compounds badly: AM1 incorrectly calculates a sharp potential barrier at a distance of approximately 3.0 Å, which may distort the calculated structures, and give erroneous activation barriers for reactions involving phosphorus [27]. PM3 shows a distinct improvement over AM1 in its treatment of phosphorus, and also considerably improved accuracy for hypervalent compounds.

2. Molecular mechanics

Molecular mechanics (MM) methods, known as force field methods, are ideal for modelling such large systems. These methods employ a potential that ignores the electron distribution of a system and instead merely considers the positions of the nuclei within the system. The resulting potential energy function known as a force field is based on the interactions between the nuclei within the system and takes into account molecular processes.

Molecular mechanics force field is constructed and parameterized by comparison with a number of molecules. This force field then can be used for other molecules similar to those for which it was parameterized. To make a molecular mechanics calculation, a force field is chosen and suitable molecular structure values (natural bond lengths, angles, etc.) are set. The structure then is optimized by changing the structure incrementally to minimize the strain energy and spread it over the entire molecule. This minimization is orders of magnitude faster than a quantum mechanical calculation on an equivalent molecule so that it is reasonable to use molecular mechanics force field instead of quantum mechanical calculation for molecular dynamics simulations. Up to date, there are a number of different MM potentials that are in common use including the CHARMM, AMBER, GROMOS and OPLS potentials. In this study, CHARMM27 force field was used, because it is the force field that was developed specifically for the simulations of organic molecules, nucleic acids, and proteins.

Molecular mechanics is a classical mechanical model that represents a molecule as a group of atoms held together by elastic bonds. Molecular mechanics looks at the bonds as springs which can be stretched called bond stretching. Compressed, bent at the bond angles, twisted in torsional angles and non-bonded interactions between atoms are also considered. The sum of all these forces is called the "force field" of the molecule. Many of the modeling force fields can be interpreted in terms of a relatively simple four component picture of the intra- and inter-molecular forces within the system. Energetic penalties are associated with the

deviation of bonds and angles away from their reference and equilibrium values, there is a function that describes how the energy changes as bonds are rotated, and finally the force field contains terms that describe the interaction between non-bonded parts of the system. More sophisticated force fields may have additional terms, but invariably contain these four components, so that it can be writing the general equation of the total energy of motion to be shown as:

$$V_{tot} \;=\; V_{bond} + V_{angle} + V_{torsion} + V_{vdw} + V_{elec} \tag{36}$$

2.1 Bonded interactions

Bonds stretching: The interaction between two atoms directly bonded to each other is assumed to be harmonic. Bond stretching energy can be represented as following,

$$E = \sum_{n=1}^{i} k_b (r - r_0)^2 \tag{37}$$

Where   r is the length of the ith bond (Å)

                   r_0 is the equilibrium length for the ith bond (Å)

                   k_b is the bond stretching constant (kcal/mol (Å)^2)

Bond angles: The interaction between three connected atoms is also assumed to be harmonic. Angle bending energy is a function of angular displacement. The bending energy equation is also based on Hooke's law

$$E = \sum_{n=1}^{i} k_\theta (\theta - \theta_0)^2 \tag{38}$$

Where θ is the angle between two adjacent bonds ($^\circ$)

                   θ_0 is the equilibrium value for the ith angle ($^\circ$)

                   k_θ is the angle bending constant (kcal/mol ($^\circ$)^2)

Dihedral (Torsion) angles: There are four atoms connected by three bonds, so that if we look straight down the second bond, the dihedral angle is the angle between the first and third bonds. The energies associated with dihedral angles are treated using a cosine series. The torsion energy is modeled by a simple periodic function.

$$E = \sum_{torsions} A[1 + Cos(n\tau - \phi)] \tag{39}$$

Where $\quad$ A is the torsional barrier (kcal/mol)

$n\tau$ is the periodicity

ϕ is the torsional angle

2.2 Non-bonded interactions

The non-bonded energy represents the pair-wise sum of the energies of all possible interacting non-bonded atoms i and j;

$$E = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}} \tag{40}$$

where the first term of the right-hand side of the equation (40) is the Lennard-Jones potential energy function that represented the van der Waals attraction and repulsion between atom i and j. The second term is the electrostatic interaction between atom i and j.

Van der Waals attraction: Van der Waals attraction occurs at short range, and rapidly dies off as the interacting atoms move apart.

Repulsion: Repulsion occurs when the distance between interacting atoms becomes even slightly less than the sum of their contact distance.

Electrostatic interactions: Electrostatic energy dies out slowly and it can affect atoms quite far apart.

3. Combined quantum mechanical/molecular mechanical methods

A key issue in the modeling of reactions in proteins is the size of the simulation system involved. A chemical reaction need to be described by the use of quantum mechanics. Although, different levels of quantum mechanical methods (for example; the semiempirical, ab initio and DFT levels of theory) can be used, the limitations of such methods are on the number of atoms that can be included in a computational model. In contrast, the classical molecular mechanical techniques which can handle whole enzyme are inefficient to explain and account for a chemical reaction. During the last decade, the combined quantum mechanical/molecular mechanical (QM/MM) approach that is a way to tackle this problem becomes increasingly popular.

Basically, the combined QM/MM methods into the description of enzymes were originally developed in the study of lysozyme by Warshel and Levitt in 1976. In general, a chemical reaction requires a quantum mechanical treatment, as electronic calculations are involved. At that time, they realized that quantum mechanical calculations were feasible only for small chemical systems and, thus, the substrate-enzyme reactions should be represented by a small modelled region. However, the outer region around the reacting parts provides mainly conformational and non-bonded contributions. These contributions can be adequately accounted by the classical molecular mechanics (MM) and the electrostatic interaction of classical particles with the reacting (QM) region. Therefore the system can be divided into two regions; (i) a small region involved in the reaction at the active site to be described quantum mechanically and (ii) the surrounding protein, the rest of system, can be adequately represented by simpler molecular mechanics. Thus, the total energy expression of the whole system (i.e. the enzyme as well as surrounding solvent) is calculated by solving the Schrödinger equation with an effective Hamiltonian for the mixed quantum and classical system:

$$\hat{H}_{\text{eff}}\Psi(r, R_\alpha, R_M) = E(R_\alpha, R_M) \tag{41}$$

where Ψ is the electronic wave function of the quantum system. It depends directly on the electron coordinates r and parametrically on the coordinates of the quantum and classical nuclei; they are referred to as R_α and R_M, respectively. The Hamiltonian can be partitioned into quantum and classical components by writing

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \tag{42}$$

where $\hat{H}_{QM}$ is the pure quantum Hamiltonian, $\hat{H}_{MM}$ is the classical Hamiltonian and $\hat{H}_{QM/MM}$ is the hybrid QM-MM Hamiltonian. Given equation (41), the total energy can be written

$$E(R_\alpha, R_M) = \frac{\left\langle \Psi \left| \hat{H}_{QM} \right| \Psi \right\rangle + \left\langle \Psi \left| \hat{H}_{QM/MM} \right| \Psi \right\rangle}{\left\langle \Psi \middle| \Psi \right\rangle}$$

$$= E_{QM} + E_{MM} + E_{QM/MM} \tag{43}$$

where E_{QM} stands for the energy of QM region, calculated with quantum mechanics; E_{MM} is the energy of MM region, calculated using an appropriate potential function; and $E_{QM/MM}$ represents the interaction energy term between the QM and the MM regions.

In this thesis, the semiempirical QM/MM methods were chosen to apply on the RT/dsDNA/dTTP ternary complex; therefore, these combined methods will be briefly presented here. According to Field et al., the electronic contribution to the interaction of the QM molecule with a point charge is determined using an expression in equation (44) based on the NDDO (neglect of differential atomic overlap) scheme adopted in semiempirical methods.

$$V_{QM}^{el}(R) = -\sum_{\mu, \nu} P_{\mu\nu}(\mu\nu | s^m s^m) \tag{44}$$

where s^m is a notional s orbital on m and μ, ν are atomic orbitals belonging to the same QM atom. The two-center two-electron integrals in equation (44) are computed in terms of the interaction of a finite multipole expansion of charges about the relevant atoms (ss, sp, and pp distributions on QM atoms are treated as monopole, dipole, and quadrupole expansions with corresponding parameters ρ^q_0, ρ^q_1 and ρ^q_2). For example, for an s-orbital distribution the multipole expansion is simply a point charge centered on the atom and the integral takes the form noted in equation (45).

$$\left(s^q s^q \middle| s^m s^m\right) = \left[R_{qm}^2 + \left(\rho_0^q + \rho_0^m\right)^2\right]^{-\frac{1}{2}} \tag{45}$$

where R_{qm} is the distance between the centers q and m, and $(\rho_0^q + \rho_0^m)^2$ is the Ohno-Kloppman factor that accounts for damping of classical Coulomb interactions due to overlapping electron densities.

The nuclear contribution follows the expression adopted for core-core interactions, as noted in equations 46-48, where α, *K, L, and M* depend on atom type and Z_q is the core charge of the q atom. The exponential terms in *f(q,m)* (eqation 47) were introduced in MNDO to correct the lack of penetration effects in the electronic component (the negative sign for the exponential terms holds for negative charges q_m so that such terms always represent a repulsive contribution). The function *g(q,m)* is an additional term in AM1 and PM3 introduced to correct deficiencies of the MNDO method in hydrogen-bonded interactions.

$$V_{QM}^{nuc}(R) = \sum_q Z_q \left[\left(s^q s^q \middle| s^m s^m\right) f(q,m) + \frac{1}{R_{qm}} g(q,m)\right] \tag{46}$$

where the sum runs over all the q atoms in the QM system.

$$f(q,m) = 1 \pm e^{-\alpha_q R_{qm}} \pm e^{-\alpha_m R_{qm}} \tag{47}$$

$$g(q,m) = \sum_i K_{q,i}\, e^{-L_{q,i}(R_{qm}-M_{q,i})^2} + \sum_j K_{m,j}\, e^{-L_{m,j}(R_{qm}-M_{m,j})^2} \tag{48}$$

Field *et al.* left unchanged the semiempirical parameters on QM atoms and optimized those on MM sites. In the final expression, $\rho_0^{\,m}$ adopted a value of zero, the terms corresponding to MM atoms in the function *g(q,m)* were omitted, and only one parameter per MM atom type, R_m, was used and set to 5.0 (au^{-1}) in order to get the best fit to ab initio data. Most alternative algorithms have continuously developed to compute $V_{QM}(R)$ related to the proceeding equations.

In CHARMM program, QM/MM electrostatic interactions for semiempirical methods are accounted by including the point charges from the MM part as atomic 'cores' which represent the nuclei and core electrons. Van der Waals interactions are treated with the molecular mechanics approach generally using a Lennard-Jones potential, consequently, MM van der Waals parameters are assigned on each QM atom. Bonded interaction terms (bonds, angles, dihedrals energies) between QM and MM atoms are retained as long as at least one MM atom involved in interaction with QM atom.

The main advantage of combined QM/MM methods is its easy implementation in computational codes giving reasonable chemical results. However, there are some problems mainly due to three factors: (i) the requirement of high level of theory in ab initio description applied for the QM region whereas the QM size would mostly allowed for semiempirical treatments; (ii) the need for accessing free energy numbers through extensive sampling which is too computationally expensive in ab initio QM/MM MD simulations; (iii) the difficult calibration of the interaction between the QM region and the MM region, especially in biochemical systems.

4. Molecular dynamics simulation

The molecular dynamics simulations method is based on Newton's second law or the equation of motion, F= ma, where F is the force exerted on the particle, m is its

mass and a is its acceleration. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Newton's equation of motion is given by

$$\mathbf{F_i} = \mathbf{m_i a_i} \tag{49}$$

Where $\mathbf{F_i}$ is the force exerted on the particle $\mathbf{i}$, $\mathbf{m_i}$ is the mass on the particle $\mathbf{i}$, and $\mathbf{a_i}$ is its acceleration. The force can also be expressed as the gradient of the potential energy.

$$\mathbf{F_i} = \nabla_{\mathbf{i}} \mathbf{V} \tag{50}$$

Combining these two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{51}$$

where V is the potential energy of the system. Newton's equation of motion can be related the derivative of the potential energy to the changes in position as a function of time.

In case of property calculations of the system,

$$\mathbf{F} = \mathbf{ma} = \mathbf{m}\frac{dv}{dt} = \mathbf{m}\frac{d^2 x}{dt^2} \tag{52}$$

and the acceleration is constant,

$$\mathbf{a} \; = \; \frac{dv}{dt} \tag{53}$$

Expression for the velocity after integration is obtained,

$$\mathbf{v} \; = \; \mathbf{a}t + \mathbf{v_0} \tag{54}$$

and since

$$\mathbf{v} \; = \; \frac{dx}{dt} \tag{55}$$

therefore,

$$\mathbf{x} \; = \; \mathbf{v}t + \mathbf{x_0} \tag{56}$$

Combining this equation for the velocity, it can be obtained the following relation which gives the value of $\mathbf{x}$ at time $\mathbf{t}$ as a function of the acceleration, $\mathbf{a}$, the initial position, $\mathbf{x_0}$ and the initial velocity, $\mathbf{v_0}$.

The acceralation is given as the derivative of the potential energy with respect to the position, $\mathbf{r}$,

$$\mathbf{a} = \; -\frac{1}{m}\frac{dE}{dr} \tag{57}$$

Therefore, to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic, e.g., the position and the velocities at time zero determine the positions and velocities

at all other times, t. The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein. It is possible to calculate the new position of each atom in the system at the end of each time step. In order to propagate a trajectory for the system of particles, the equation (57) must be integrated for all particles in the system at each timestep of simulation. This is not trivial task because the motions of all atoms in the system are coupled together and thus analytically solving the equations of motion is impossible. Therefore, these equations must be integrated using numerical methods.

5. Algorithm of molecular dynamics simulations

The potential energy is a function of the atomic position (3N) of all the atoms in the system. Due to the complicated nature of this function, it must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion as shown in equation (58) – (60):

$$\mathbf{r(t+\Delta t)} \ = \ \mathbf{r(t)} + \mathbf{v(t)}(t + (1/2)a(t)(t2 + ...} \qquad (58)$$

$$v(t+(t) \ = \ v(t) + a(t)(t + (1/2)b(t)(t2 + ... \qquad (59)$$

$$a(t+(t) \ = \ a(t) + b(t)(t + ... \qquad (60)$$

where **r** is the position, **v** is the velocity (the first derivative with respect to time), **a** is the acceleration (the second derivative with respect to time), and etc. Mostly MD software packages, including CHARMM, use a variety of methods based on the Verlet algorithm for propagation of atomic coordinates, which are designed to allow velocities to be calculated at each step.

The basic idea of Verlet algorithm (Verlet, 1967) is to write two third-order Taylor expansions for the positions **r(t)**, one forward and one backward in time.

Calling **v** the velocities, **a** the accelerations, and b the third derivatives of r with respect to t, as following:

$$\mathbf{r(t+\Delta t) = r(t) + v(t)}(t + (1/2)a(t)(t2 + (1/6)b(t)(t3 + O((t4) \tag{61}$$

$$r(t-(t) = r(t) - v(t)(t + (1/2)a(t)(t2 - (1/6)b(t)(t3 + O((t4) \tag{62}$$

Adding the two expressions gives

$$\mathbf{r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)}(t2 + O((t4) \tag{63}$$

The equation (63) is the basic form of the Verlet algorithm. Since the integrating Newton's equation, **a(t)** is just the force divided by the mass and the force is in turn a function of the position **r(t)**,

$$\mathbf{a(t) = (1/m)....V(r(t))} \tag{64}$$

As one can immediately see, the truncation error of the algorithm when evolving the system by **Δt** is of the order of $\mathbf{\Delta t^4}$, even if third derivatives do not appear explicitly. This algorithm is at the same time simple to implement accurate and stable, explaining its large popularity among molecular dynamics simulations.

A problem of the Verlet algorithm is that velocities are not directly generated. While they are not needed for the time evolution, their knowledge is sometimes necessary. Moreover, they are required to compute the kinetic energy **K**, whose evaluation is necessary to test the conservation of the total energy **E** = **K+V**. This is one of the most important tests to verify that a MD simulation is proceeding correctly. One could compute the velocities from the positions by using equation (65).

$$\mathbf{v(t)} = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} \tag{65}$$

However, the error associated to this expression is of order Δt^2 rather than Δt^4.

An algorithm always used in molecular dynamics simulation is the SHAKE algorithm which applies constrains to the system in an iterative manner until all constrains satisfied. The SHAKE algorithm is applied to fix all MM bonds involving hydrogen atoms. Consequently, when used in conjunction with the Verlet algorithm, the SHAKE algorithm has been found to reduce computational expense significantly. The procedure is based on the determination of Lagrange multipliers (l_k) imposed as a constriction to the equations of motion. For n constraints with d_k as the corresponding constrain distance, we have

$$m_i \frac{d^2 q_i(t)}{dt^2} = -\frac{\delta}{\delta q_i}\left[V(q)+\sum_k^n l_k(t)(q_k^2 - d_k^2)\right] \tag{66}$$

To solve the equation (66) in SHAKE procedure, the Lagrange multipliers are determined iteratively.

6. Umbrella sampling method

In the umbrella sampling method, the microscopic system is simulated in the presence of an artificial biasing window potential $V_b(\chi)$ that is added to the potential energy $V(q)$.

$$V' = V(q) + V_b(\chi) \tag{67}$$

The artificial biasing window potential $V_b(\chi)$ was applied on system for restrained dynamics simulation within a small interval window of a prescribed value of χ. Although, the entire range of χ coordinate is spanned in a single simulation (window), multiple simulations (windows) are performed with different biasing umbrella potentials $V_b(\chi)_i$ that center the sampling in different overlapping regions or windows of χ. For the ith simulation the artificial bias window potential was given as a harmonic potential by

$$V_b(\chi)_i = \frac{1}{2}k(\chi - \chi_i)^2 \tag{68}$$

where the value of the energy constant k in this thesis was chosen to be 200 and 300 kcal mol^{-1} with χ coordinate in Å for the reactions of deprotonation and DNA polymerization, respectively.

The Weighted Histogram Analysis Method (WHAM) computes the total unbiased distribution function $\langle\rho(\chi)\rangle^u$ as a weighted sum of the unbiased distribution functions $\langle\rho(\chi)\rangle_i^u$. This weighting function can be expressed in terms of the known biased distribution functions $\langle\rho(\chi)\rangle_i^b$.

$$\langle\rho(\chi)\rangle^u = \sum_{i=1}^{N} n_i \langle\rho(\chi)_i^b\rangle \times \left[\sum_{j=1}^{N} n_j e^{(f_i - V_b(\chi)i)/k_B T} \right]^{-1} \tag{69}$$

where N is the total number of windows and n_i is the number of data points at χ (number of states in a MD simulation) in the window. The free energy constants f_i need to be obtained from an optimal estimate of the total distribution function

$$e^{-f_i/k_B T} = \int d\chi\, e^{V_b(\chi)_i/k_B T} \langle\rho(\chi)\rangle^u \tag{70}$$

Again, $V_b(\chi)_i$ is the bias potential at χ coordinate for the ith simulation. Since the window free energy constants depend on the probability and the probability depends on the window free energy constants, equations 69 and 70 need to be solved iteratively until self-consistence. A common procedure is giving a guess to the set of free energy constants $\{f_i\}$, obtain the total distribution function by equation 69 and use it to compute $\{f_i\}$ in equation 70. The procedure is stopped when the difference between the constants in two consecutive iterations is below a threshold, generally within 0.1%.

APPENDIX B: Presentations and Proceedings

Oral Presentation and Proceedings

1. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Modelling the enzymatic reaction of HIV-1 RT based on combined QM/MM approach. In the proceeding of Workshop on Modeling Interactions in Biomolecules II, Czech University of Agriculture, Prague, Czech of Replubic, 5-9 Sep 2005, page 521.

2. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Hybrid QM/MM study on the active site of HIV-1 RT. In the proceeding of Royal Golden Jubilee Ph.D. Congress VI, Jomtien Palm Beach Resort Patthaya, Chonburi, Thailand, 28-29 April 2005, page 114.

3. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Mechanistic investigation on the active site of HIV-1 RT based on QM/MM umbrella sampling method. The 8th Annual National Symposium on Computational Science and Engineering (ANSCSE8), Suranaree University of Technology, Nakhon Ratchasima, Thailand, 21-23 July 2004.

4. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Model of Mg^{2+} complex with the HIV-1 RT based on ONIOM method Mechanistic. The 7th Annual National Symposium on Computational Science and Engineering (ANSCSE7), Chulalongkorn University, Bangkok, Thailand, 24-26 March 2003.

Poster Contribution to Conferences

1. Suwipa Saen-oon, Mayuso Kuno, Thanyada Rungrotmongkol, Perapol Nunrium and Supa Hannongbua. Drug-enzyme interaction; Structural information for inhibitor design. In the proceeding of XIIth International Congress of Computational Chemistry (XIIth ICQC), Kyoto Terrsa, Kyoto, Japan, 21-26 May 2006, page 114.

2. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Molecular mechanics and combined quantum mechanical and molecular mechanical (QM/MM) simulations on enzymatic reaction of HIV-1 RT active. 2nd Asian Pacific Conference on Theoretical and Computational Chemistry (APCTCC2), Chulalongkorn University, Bangkok, Thailand, 2-6 May 2005.

3. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Mechanistic study of HIV-1 Reverse Transcriptase at the active site based on QM/MM method. 1st Asian Pacific Conference on Theoretical and Computational Chemistry (APCTCC1), Institute for Molecular Science (IMS), Okazaki, Japan, 12-15 May 2004.

4. Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Modelling the structure and mechanism of HIV-1 reverse transcriptase. MGMS Young Modellers' Forum in Conjunction with the RSC MMG, London, U.K., 1 Dec 2003.

Thanyada Rungrotmongkol

Laboratory for Computational and Applied Chemistry

Department of Chemistry, Faculty of Science, Kasetsart University

50 Paholyothin Rd. Chatuchak, Bangkok, THAILAND 10900.

PERSONAL DETAILS

Date of Birth: 14th January 1979

Place of Birth: Rajburi, Thailand

Nationality: Thai

Marital Status: Single

Age: 27

Phone Number: +66-2942-8900 ext.325

Fax Number: +66-2942-8900 ext.324

E-mail address: g4484013@ku.ac.th, T.Rungrotmongkol@gmail.com

EDUCATION

Kasetsart University, Bangkok, Thailand

- Ph.D. in Physical Chemistry   2001-Present
- B.Sc. in Chemistry (1st honour)   1997-2000

RESEARCH VISITING

School of Chemistry, University of Bristol, UK

- Research work with Dr. Adrian J. Mulholland   Jul. 2003 – Mar. 2004
- Research work with Dr. Adrian J. Mulholland   Nov. 2005 – Jan. 2006

ACADEMICAL EXPERIENCE

Research Assistant (2001- present)

Laboratory for Computational and Applied Chemistry Physical Chemistry Division,

Kasetsart University, Bangkok, Thailand

RESEARCH INTERESTS

- Structure and catalytic reactivity of HIV-1 Reverse Transcriptase
- Reaction mechanism of DNA polymerization on active site of HIV-1 RT
- Molecular dynamics simulation in biomolecular systems
- Particular interactions between inhibitor and its binding pocket

PUBLICATIONS

1. **Mechanistic study of HIV-1 Reverse Transcriptase at the active site based on QM/MM method.**
   Thanyada Rungrotmongkol, Supa Hannongbua and Adrian J. Mulholland.
   Journal of Theoretical and Computational Chemistry **(2004)**, 3 (4), 491-500.

2. **Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modeling of a HIV-1 Reverse Transcriptase/DNA/dTTP complex.**
   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua.
   Submitted to Journal of molecular graphics and modelling, accepted

3. **Combined QM/MM modelling of HIV-1 RT: Discriminating between alternative mechanisms in DNA polymerization.**
   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. in preparation.

CONFERENCES (ORAL AND POSTER PRESENTATION)

1. **Drug-enzyme interaction; Structural information for inhibitor design**.

Suwipa Saen-oon, Mayuso Kuno, Thanyada Rungrotmongkol, Perapol Nunrium and Supa Hannongbua. XIIth International Congress of Computational Chemistry (XIIth ICQC), Kyoto Terrsa, Kyoto, Japan, May 21-26 **2006.** (Poster presentation)

2. **Modelling the enzymatic reaction of HIV-1 RT based on combined QM/MM approach.**

   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. Workshop on Modeling Interactions in Biomolecules II, Czech University of Agriculture, Prague, Czech, September 5-9 **2005**. (Oral Presentation)

3. **Molecular mechanics and combined quantum mechanical and molecular mechanical (QM/MM) simulations on enzymatic reaction of HIV-1 RT active.**

   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua, 2nd Asian Pacific Conference on Theoretical and Computational Chemistry (APCTCC2), Chulalongkorn University, Bangkok, Thailand, May 2-6 **2005.** (Poster presentation)

4. **Hybrid QM/MM study on the active site of HIV-1 RT**.

   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. The Royal Golden Jubilee Ph.D. Congress VI, Pattaya, Thailand, April 28-29 **2005**. (Oral Presentation)

5. **Mechanistic investigation on the active site of HIV-1 RT based on QM/MM umbrella sampling method**.

   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. The 8th Annual National Symposium on Computational Science and Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand, July 21-23 **2004**. (Oral Presentation)

6. **Mechanistic study of HIV-1 Reverse Transcriptase at the active site based on QM/MM method**.

   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. 1st Asian Pacific Conference on Theoretical and Computational Chemistry (APCTCC1), Institute for Molecular Science (IMS), Okazaki, Japan, May 12-15 **2004**. (Poster presentation)

7. **Modelling the structure and mechanism of HIV-1 reverse transcriptase.**
   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. MGMS Young Modellers' Forum in Conjunction with the RSC MMG, London, U.K., December 1 **2003**. (Poster presentation)

8. **Model of Mg^{2+} complex with the HIV-1 RT based on ONIOM method**.
   Thanyada Rungrotmongkol, Adrian J. Mulholland and Supa Hannongbua. The 7th Annual National Symposium on Computational Science and Engineering, Chulalongkorn University, Bangkok, Thailand, March 24-26 **2003**. (Oral Presentation)

9. **Structural Calculations of Dihydrofolate Reductase Inhibitor, Cycloguanil.**
   Thanyada Rungrotmongkol and Supa Hannongbua. The International Conference on Bioinformatics 2002: North-South Networking (INCOB 2002), Le Royal Meridien, Bangkok, Thailand, February 6-8 **2002**.

SCHOLARSHIPS AND AWARDS

1. Oral presentation award of The Royal Golden Jubilee Ph.D. Congress VI. **(2005)**

2. The Royal Golden Jubilee Ph.D. Program, Thailand Research Fund. **(2001-present)**

3. Postgraduate Education and Research Program in Physical Chemistry, Higher Education Development Project Fund. **(2001-present)**

4. Research Assistant Scholarship. **(2001-present)**

5. Top score award in Department of Chemistry, Faculty of Science, Kasetsart University from Tabb Neratit Foundation, Thailand **(2001)**

6. Top score award in Department of Chemistry, Faculty of Science, Kasetsart University from The Siam Cement Group, Thailand **(2000)**

7. Top score award in Department of Chemistry, Faculty of Science, Kasetsart University from Tabb Neratit Foundation, Thailand **(1998)**

8. Development and Promotion of Science and Technology talents project of Thailand (DPST) **(1997-2000)**