Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 1–13
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex

Thanyada Rungrotmongkol [a], Adrian J. Mulholland [b], Supa Hannongbua [a,*]

[a] *Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand*
[b] *Computational Chemistry Centre, School of Chemistry, University of Bristol, Bristol, UK*

Received 29 June 2006; received in revised form 7 September 2006; accepted 10 September 2006
Available online 14 September 2006

Abstract

We have investigated the structure and dynamics of the HIV-1 reverse transcriptase (HIV-RT) active site, by modelling the active conformation of the HIV-1 RT/DNA/deoxythymidine triphosphate (dTTP) ternary complex. This has included molecular dynamics simulations with the CHARMM27 force field, and combined quantum mechanics/molecular mechanics (QM/MM) calculations. Three different ternary systems were studied to investigate the effects of different protonation states of the dTTP substrate (a deprotonated and two different mono-protonated triphosphate forms of dTTP at the active site), and the effects of different possible protonation state of potentially catalytic aspartate residues (Asp185 and Asp186) were tested. Several potentially important hydrogen-bonding interactions with amino acids and bound water molecules in the deoxyribonucleoside triphosphate (dNTP) binding pocket were examined. The model of the deprotonated form of the dTTP substrate with the two aspartates in their charged (basic) form seemed to be the most stable and its orientation was in good agreement with X-ray crystallographic structure. In addition, two different semiempirical (AM1/CHARMM and PM3/CHARMM) QM/MM methods were tested for the HIV-RT system, in structural optimizations. Both methods provided conformations of the triphosphate moiety in either fully deprotonated or mono-protonated forms, which agreed well with the experimental structure of dTTP. The only significant difference between the AM1/CHARMM and PM3/CHARMM minimized structures is that the PM3/CHARMM Pα–O3′ optimized distance (important for nucleotide addition) is longer by 0.66 Å in the deprotonated system but shorter by 0.37 Å in the mono-protonated triphosphate system as compared with those obtained from AM1/CHARMM minimized structure. The obtained results suggest that both of these QM/MM methods, and the stochastic boundary molecular dynamics approach applied in this work, can give reasonable results for modelling the catalytically active complex of this important enzyme. The results provide insight into the structure and interactions of the active site of this important enzyme, with implications for its mechanism, which may be useful in inhibitor design.

Keywords: HIV-1 RT; AIDS; Molecular dynamics; QM/MM; Semiempirical method

1. Introduction

Reverse transcriptase (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 [1]. 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' and connection subdomains, with the RNaseH domain found only in the p66 subunit [2]. 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 DNA-binding cleft formed by the p66 fingers, palm and thumb subdomains. 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 [3–7]. DNA polymerization in polymerases is thought to

* Corresponding author. Tel.: +66 2 9428900x217; fax: +66 2 9428900x218.
E-mail address: fscisph@ku.ac.th (S. Hannongbua).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.09.004

involve a two metal ion mechanism [8,9], 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 Fig. 1). 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.

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 classes: nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs, such as AZT, ddC and ddI [10,11], 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. In contrast, NNRTIs, e.g. nevirapine, TIBO, HEPT and Efavirenz [12,13], are non-competitive inhibitors which bind at a common allosteric site. It is thought that these act by distorting the HIV-1 RT enzyme in several important regions near the polymerase active site, forcing it to adopt an inactive conformation [14]. NNRTIs also affect the regions that contact the nucleic acid [12,13]. 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 [15–17] 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 [18] 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

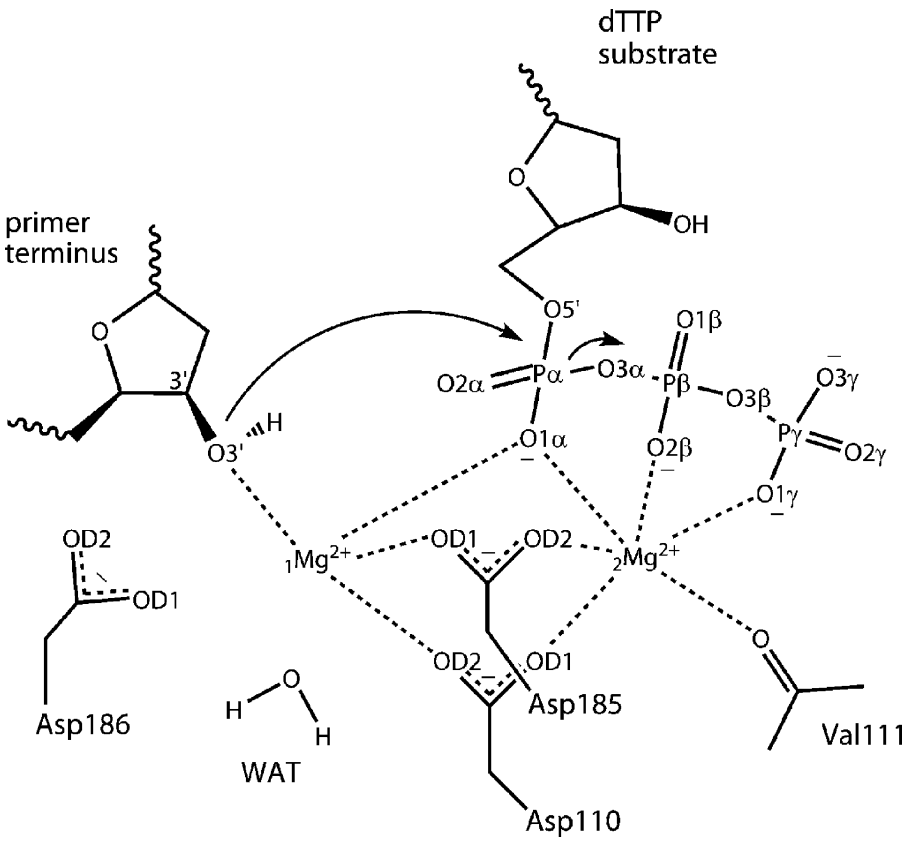

Fig. 1. The proposed mechanism [8,9] of DNA polymerization of HIV-1 RT. The 3′-hydroxyl group of the primer terminus attacks the α-phosphate of the dTTP substrate, forming a stable pentacoordinate intermediate, then giving the 3′–5′ phosphate diester and pyrophosphate as the final products. Prior to nucleotide addition, the 3′-hydroxyl hydrogen on the DNA primer terminus should transfer to a H-accepter. Coordination of both magnesium ions is shown as dashed lines, based on the crystal structure of the HIV-1 RT/DNA/dTTP ternary complex [36]. This is Model A (dTTP deprotonated). Models B and C have dTTP mono-protonated on the O2γ or O3γ atom, respectively.

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.

Combined quantum mechanical/molecular mechanical (QM/MM) methods make possible the modelling of chemical reactions in large systems such as enzymes [19,20]. In QM/MM calculations, the reactive centre (i.e. a small region such as the enzyme active site) is treated quantum mechanically and the rest of the system (the surrounding protein and solvent) is considered classically by molecular mechanics. QM/MM methods have been used to study many enzyme–catalysed reactions [21,22] which the basic idea was originated by Warshel and Levitt for the study of the catalytic mechanism of lysozyme [23].

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) [22] 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 [24–27] to reduce the computational time. It is possible to carry out QM/MM calculations with *ab initio* [27,28] or density-functional theory (DFT) [27–29] QM approaches, but these are more expensive when the QM subsystem has many atoms, and prohibitively expensive when molecular dynamics simulations are required.

In previous preliminary work, we used MM modelling (molecular dynamics simulations with the CHARMM27 force field) to study binding interactions at the active site and the conformational behavior of the enzyme, and QM/MM (AM1/CHARMM) methods to investigate the deprotonation and DNA polymerization reactions [30]. The methyltriphosphate of substrate complexed with two Mg^{2+} ions, the methanol portion of deoxyribose ring of the 3′-terminal DNA primer and the carboxylate sidechain of D185 were described quantum mechanically, whereas other atoms were treated molecular mechanically (CHARMM27). D185 was selected to be the proton acceptor in deprotonation. The results indicated that it abstracts the 3′-OH hydrogen atom from the primer terminus prior to the nucleophilic attack. A pentacoordinate structure was found to be an intermediate. In the product complex, the QM/MM results showed that the nucleotide-binding magnesium ion ($_2Mg^{2+}$) stabilizes the pyrophosphate leaving group as it moves away from the newly formed 3′–5′ phosphate diester, while $_1Mg^{2+}$ maintains its position bridging this 3′–5′ phosphate diester. However, the free energy profile showed that the final product is unstable compared to the intermediate and reactant complexes. Many questions remain unanswered and the detailed reaction pathway is still a matter of debate. There is therefore a need to compare alternative mechanistic ideas (e.g. D185 or D186 as the base, and the possibility of different protonation states of the dTTP substrate, e.g. whether it exists in a deprotonated, mono-protonated or di-protonated form). Computational methods provide a good means of investigating such questions. DNA polymerization has also been modelled in other polymerases such as T7 DNA polymerase [31] and DNA polymerase β [32–34]. In the catalytic reaction of T7 DNA polymerase, free energy perturbation/empirical valence bond (FEP/EVB) calculations indicated the most favorable reaction mechanism involved proton transfer to D654 prior to nucleophilic attack, then, followed by Pα–O3α bond cleavage. The mechanistic reactions in polymerase β were studied by calculations on active site models using density-functional theory: the results showed that the mechanism via the pentacoordinate intermediate appeared to be feasible. While an α-phosphate oxygen atom acted as H-acceptor, the protonated D192 had a hydrogen-bonding interaction with α-phosphate. Whether these enzymes also have three equivalent residues in the active site (the D654–D475–E655 triad: T7 DNA polymerase and the D192–D190–D256 triad: DNA polymerase β), and to the D185–D110–D186 triad in RT, they are equivalent.

Here, MM molecular dynamics simulations, and QM/MM optimizations have been applied to model the ternary complex of HIV-1 RT/DNA/deoxythymidine triphosphate (dTTP) substrate (Fig. 1). The main objective of the present study was to investigate the structural and dynamical properties of both protonation states of the dTTP substrate in the HIV-1 RT active site, and to test different QM/MM methods (AM1/CHARMM and PM3/CHARMM). 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 [35]. PM3 shows a distinct improvement over AM1 in its treatment of phosphorus, and also considerably improved accuracy for hypervalent compounds.

2. Methods

2.1. Structure of HIV-1 RT

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/DNA/dTTP ternary complex, whose crystal structure has been solved at 3.2 Å resolution (PDB entry code 1RTD.PDB) [36]. 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 [37]. In non-metalloenzymes, aspartic acid residues 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 [38,39]. 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. This will be explained in more detail in the next section.

In the nucleotide-binding site, 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 ($_1Mg^{2+}$). 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. Similar coordination spheres have also been observed in high-resolution structures of ternary complexes of the enzymes DNA polymerase β [40] and T7 DNA polymerase [4], 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 (catalysed the phosphoryl transfer reaction) as mentioned in Section 1. 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 [41]. Other amino acids close to the triphosphate moiety of dTTP are G112, D113, A114, Y115 and Q151 [42,43].

2.2. 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 O2γ-oxygen (Model B) or the O3γ-oxygen (Model C). It was considered unlikely that a proton could be bound to Oγ1-oxygen which is partially coordinated to $_2Mg^{2+}$. 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 monoprotonated (Model B) forms of the triphosphate moiety. These models are denoted HD185-Model A; HD185-Model B; HD186-Model A; 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 [30]. 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 ($_1Mg^{2+}$)) 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 [44]. 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 [45]. Water molecules' behavior was influenced by the deformable boundary potential [46] 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.

2.3. Molecular dynamics simulations

All simulations were performed with the CHARMM [47,48] program, version c27b2. For all simulations with purely MM methods, a cut-off of 13 Å was applied to non-bonded interactions and the CHARMM27 parameter set [49–52] 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 25 Å of the centre were included. Hydrogen atoms were added on the basis of CHARMM internal coordinates with the HBUILD facility [53], and were minimized with two 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 [54]. 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 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 [44]. 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 [46]. 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 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 and 25 Å. 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 [45]. The outer, reservoir zone, (containing all atoms further than 25 Å from the centre) was not included in the calculations. 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 [45,55]. Therefore, two arginine, six lysine, an aspartate and seven 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 to 300 K in three 10 ps intervals: first from 10 to 100 K, then from 100 to 200 K, and, finally, from 200 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 Models A, B and C (i.e. using different initial velocities). Five hundred picosecond simulations (following 200 ps 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 [56,57] 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.

2.4. QM/MM calculations

Two different semiempirical methods (AM1 and PM3) [58–62] were applied for the QM region while the remainder of the system was treated by the MM (CHARMM27) [49–52] force field. The QM region was made up from several components as shown in Fig. 1. 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 D110, D185 and D186, the backbone carbonyl of V111 terminated with the backbone methyl-amino of G112, and a modelled water molecule were described quantum mechanically, using hydrogen 'link atoms' [63] to model the covalent bonds between carbon atoms across the QM/MM boundary. Altogether, the QM region included 58 atoms with an overall charge of −1 in Model A, and 59 atoms with neutral charge in Model B (including eight link atoms in each case). 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.

3. Results and discussion

3.1. 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 Fig. 2. 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 3×500 ps 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 amino acid residues averaged over the 3×500 ps simulations for two systems studied (Models A and B) is shown in Fig. 3(a). The results indicate that the most flexible region correspond to amino acids 64–71 (see Fig. 3(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 (Fig. 3(b)) (e.g. the average RMSD of Asp218 is ca. 0.58 Å: Model A and ca. 1.05 Å: Model B). The fully deprotonated dTTP substrate model altogether shows less movement, suggesting a more stable active site region.

Fig. 4 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

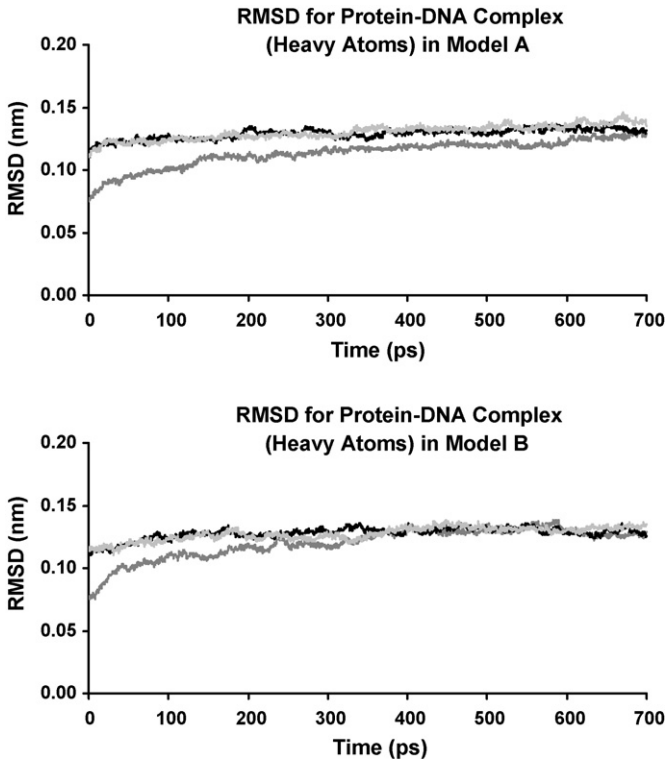

Fig. 2. RMSDs (from the initial structure) of all heavy atoms as a function of simulation time, for two systems studied here: (1) Model A (deprotonated dTTP) and (2) Model B (mono-protonated dTTP on the O2γ-oxygen). Three 700 ps MD simulations are shown in the plot with different shades.

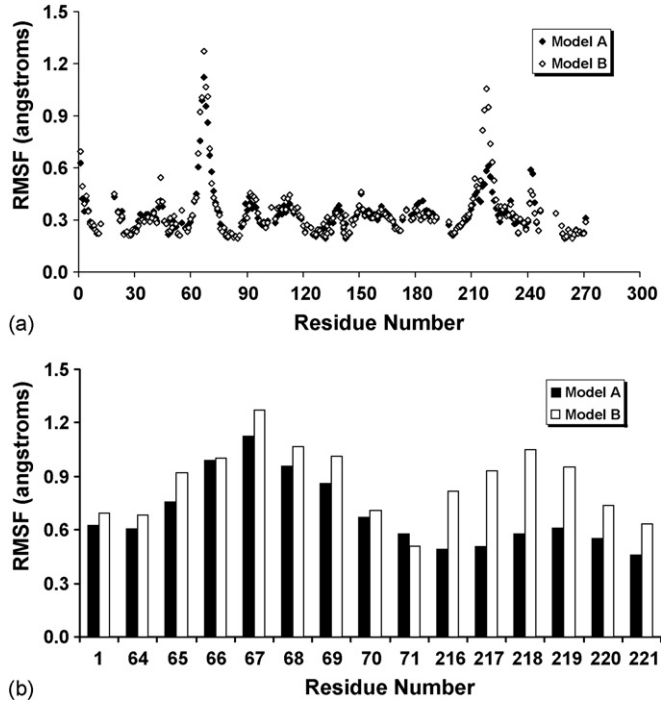

Fig. 3. Root mean square fluctuations (RMSF) (Å) of protein backbone atoms averaged over 3 × 500 ps MD simulations from two systems shown in different shades: (1) Model A (deprotonated dTTP) in black and (2) Model B (mono-protonated dTTP on the O2γ-oxygen) in white. RMSF for all residues is shown in (a) and for some selected residues in (b).

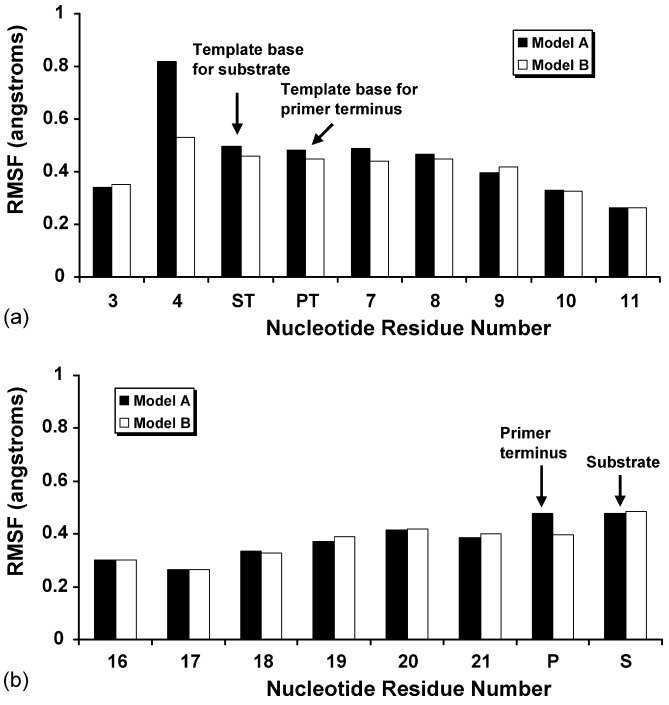

Fig. 4. Root mean square fluctuations (RMSF) (Å) of all heavy atoms by nucleotide averaged over 3 × 500 ps MD simulations for the DNA template (a), DNA primer and dTTP substrate (S) (b) from the two simulations shown in different shades: (1) Model A in black and (2) Model B in white. The template base for substrate (ST) and primer terminus (PT) are Ade5 and Cyt6 while the primer terminus (P) is Gua22.

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 1 compares the average RMSD of amino acids D110–Y115 and M184–D186, the part of the polymerase active site. 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

Table 1
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) and (2) Model B (mono-protonated dTTP on the O2γ-oxygen) with respect to the crystal structure of the HIV-1 RT/DNA/dTTP ternary complex[a] [36]

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.

(<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.

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

The average values of some important structural parameters (averaged over the 3 × 500 ps simulations) in Models A and B are shown in Table S1 (see also Fig. 1). 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 Models 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 non-bridging 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 Models 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.

In contrast, Model C, mono-protonated on the O3γ atom, showed a distorted triphosphate geometry of dTTP (Fig. 5). 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 Pβ–Pγ (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.

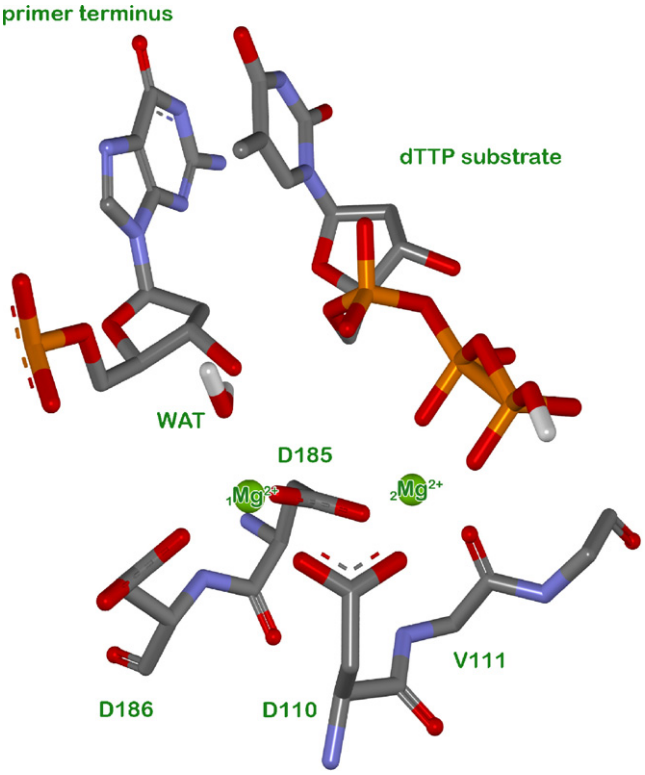

Fig. 5. Conformation of mono-protonated triphosphate on O3γ oxygen of dTTP substrate, two Mg^{2+} ions and their coordination environments (Model C) in the final structure from MM MD simulation. Hydrogen atoms are not shown except for the mono-protonated O3γ atom and the bound water molecules. The colors were used by atom type color.

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 S1, 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 Models 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 and 2.6 Å, respectively), than in Model A (where they are 2.4 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 3 × 500 ps molecular dynamics trajectories of Models 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 S1.

To analyse 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 S2) 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 S2. 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 S2: 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 [36]. 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 [64] 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.

3.3. Protonation states of the aspartic sidechain

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

Table 2 showed the active site atom distances averaged over the last 500 ps MD trajectories of four different systems.

Table 2
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 ($\angle$) 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.

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 Models A and B simulations (3.6 Å: Model A and 3.5 Å: Model B) (see Table 2 and Table S1). 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 Models 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 2. 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 2). 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 2). 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 2 and Table S1).

Table 3
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 [36]

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

Table 3 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.

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.

3.4. 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 cut-off was applied for non-bonded 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 Section 2. The starting structure for these QM/MM calculations was obtained from the MM minimized snapshot taken from the final MD structure. Table 4 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 4 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 4). 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 O2γ-oxygen), which agreed well with the crystal structure of dTTP, as shown in Fig. 6(a) and (b). The binuclear magnesium coordination environment differs between these two semiempirical methods, as shown by the Mg–O distances in Table 4. 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[Mg2–O1γ]): Model B). With AM1/CHARMM, the Mg1–O1α distance decreased from the minimized MM structures (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 the Mg1–O1α 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 ($_1Mg^{2+}$) further displaced from the crystal structure (3.57 Å) as shown in Fig. 6(a).

As can be seen in Fig. 6, 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 Models A and B. They shared this feature with their minimized MM (CHARMM) structure (see Table 4). 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

Table 4
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

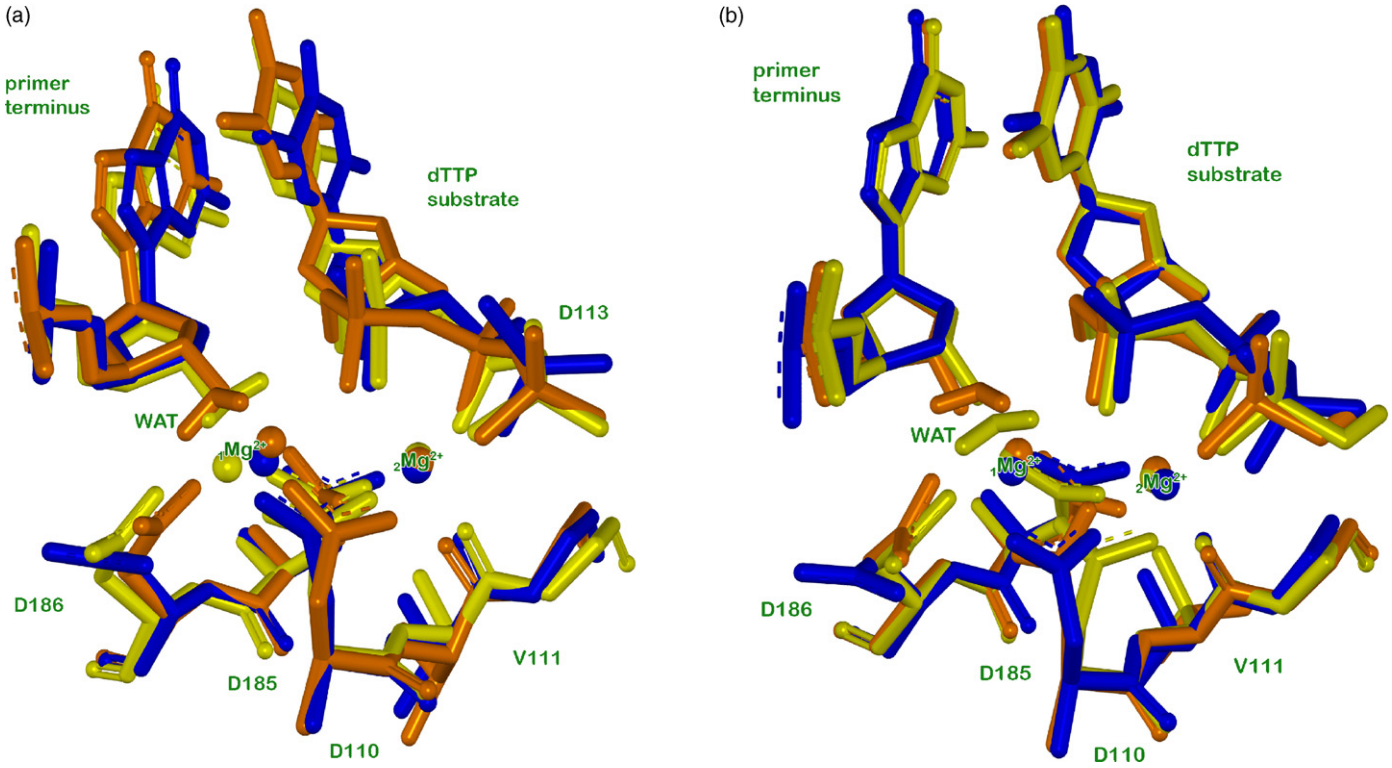

Fig. 6. Superimposition of QM/MM minimized AM1/CHARMM (orange) and PM3/CHARMM (yellow) structures (from the final MD structure) compared to the crystal structure (blue). Model A is shown in (a) and Model B in (b). Hydrogen atoms are not shown except for the mono-protonated O2γ atom (Model B) and a bound water molecule.

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 [45,65].

4. Conclusions

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 (D185 and D186). 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 K65, R72, D113 and A114 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 (D110, D185 and D186) and a modelled water molecule. The ligands of the other (bridging) magnesium ($_2Mg^{2+}$) are three phosphate oxygens, the sidechains of two bridging aspartates (D110 and D185) and a backbone carbonyl of valine (V111).

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.

Acknowledgements

This investigation was supported by the Thailand Research Fund (BRG4780007). T.R. thanks the Royal Golden Jubilee Ph.D. Program (3.C.KU/44/B.1) for scholarship. A.J.M. thanks BBSRC, EPSRC, the Wolfson Trust, and the IBM High Performance Computing Life Sciences Outreach Programme for support. The Postgraduate Education and Research on Petroleum and Petrochemical Technology, Laboratory of Computational and Applied Chemistry, Kasetsart University Research and Development Institute and the University of Bristol Computational Chemistry Centre provided research facilities, software packages and computing time.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.09.004.

References

[1] E.M. Wondrak, J. Loewer, R. Kurth, Functional purification and enzymic characterization of the RNA dependent DNA polymerase of human immunodeficiency virus, J. Gen. Virol. 67 (1986) 2791–2797.
[2] Z. Hostomsky, Z. Hostomska, T.B. Fu, J. Taylor, Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis, J. Virol. 66 (1992) 3179–3182.
[3] T.A. Steitz, DNA polymerases: structural diversity and common mechanisms, J. Biol. Chem. 274 (1999) 17395–17398.
[4] S. Doublie, S. Tabor, A.M. Long, C.C. Richardson, T. Ellenberger, Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution, Nature 391 (1998) 251–258.
[5] L.S. Beese, V. Derbyshire, T.A. Steitz, Structure of DNA polymerase I Klenow fragment bound to duplex DNA, Science 260 (1993) 352–355.
[6] J. Wang, A.K. Sattar, C.C. Wang, J.D. Karam, W.H. Konigsberg, T.A. Steitz, Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69, Cell 89 (1997) 1087–1099.
[7] A. Bibillo, D. Lener, G.J. Klarmann, S.F.J. Le Grice, Functional roles of carboxylate residues comprising the DNA polymerase active site triad of Ty3 reverse transcriptase, Nucleic Acids Res. 33 (2005) 171–181.
[8] L.S. Beese, T.A. Steitz, Structural basis for the 39–59 exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism, EMBO J. 10 (1991) 25–33.
[9] V. Derbyshire, N.D.F. Grindley, C.M. Joyce, 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 (1991) 17–24.
[10] Y. Chong, K. Borroto-Esoda, A.F. Phillip, F.S. Raymond, K.C. Chung, Molecular mechanism of DApd/DXG against zidovudine- and lamivudine-drug resistant mutants: a molecular modelling approach, Antivir. Chem. Chemother. 13 (2002) 115–128.
[11] P.S. Kedar, J. Abbotts, T. Kovacs, K. Lesiak, P. Torrence, S.H. Wilson, Mechanism of HIV reverse transcriptase: enzyme–primer interaction as revealed through studies of a dNTP analogue, 3′-azido-dTTP, Biochemistry 29 (1990) 3603–3611.
[12] S. Hannongbua, K. Nivesanond, L. Lawtrakul, P. Pungpo, P. Wolschann, 3D-quantitative structure–activity relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on ab initio calculations, J. Chem. Inf. Comput. Sci. 41 (2001) 848–855.
[13] S. Saen-Oon, S. Hannongbua, P. Wolschann, 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. Comput. Sci. 43 (2003) 1412–1422.
[14] L. Lawtrakul, A. Beyer, S. Hannongbua, P. Wolschann, Quantitative structural rearrangement of HIV-1 reverse transcriptase on binding to non-nucleoside inhibitors, Monatsh. Chem. 135 (2004) 1033–1046.
[15] G.J. Klarmann, R.A. Smith, R.F. Schinazi, T.W. North, B.D. Preston, 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 (2000) 359–366.
[16] H.Q. Gao, P.L. Boyer, S.G. Sarafianos, E. Arnold, S.H. Hughes, The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase, J. Mol. Biol. 300 (2000) 403–418.
[17] N. Sluis-Cremer, D. Arion, N. Kaushik, H. Lim, M.A. Parniak, Mutational analysis of Lys65 of HIV-1 reverse transcriptase, Biochem. J. 348 (2000) 77–82.
[18] P.R. Meyer, S.E. Matsuura, A.G. So, W.A. Scott, Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 13471–13476.
[19] L. Ridder, A.J. Mulholland, Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity, Curr. Top. Med. Chem. 3 (2003) 1241–1256.
[20] A.J. Mulholland, Modelling enzyme reaction mechanisms, specificity and catalysis, Drug Discov. Today 10 (2005) 1393–1402.
[21] B.A. Gregersen, X. Lopez, D.M. York, Hybrid QM/MM study of thio effects in transphosphorylation reactions, J. Am. Chem. Soc. 125 (2003) 7178–7179.
[22] J.C. Hermann, C. Hensen, L. Ridder, A.J. Mulholland, H.-D. Hoeltje, Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A b-lactamase with benzylpenicillin, J. Am. Chem. Soc. 127 (2005) 4454–4465.
[23] A. Warshel, M. Levitt, Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. 103 (1976) 227–249.
[24] A. Van Der Vaart, V. Gogonea, S.L. Dixon, K.M. Merz Jr., Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method, J. Comput. Chem. 21 (2000) 1494–1504.
[25] J. Khandogin, D.M. York, Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods, Proteins 56 (2004) 724–737.
[26] L. Ridder, I.M.C.M. Rietjens, J. Vervoort, A.J. Mulholland, 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 (2002) 9926–9936.
[27] D.W. Sheppard, N.A. Burton, I.H. Hillier, Ab initio hybrid quantum mechanical/molecular mechanical studies of the mechanisms of the enzymes protein kinase and thymidine phosphorylase, Theochem: J. Mol. Struct. 506 (2000) 35–44.
[28] P.D. Lyne, M. Hodoscek, M. Karplus, A hybrid QM–MM potential employing Hartree–Fock or density functional methods in the quantum region, J. Phys. Chem. A 103 (1999) 3462–3471.
[29] R.V. Stanton, D.S. Hartsough, K.M. Merz Jr., An examination of a density functional/molecular mechanical coupled potential, J. Comput. Chem. 16 (1995) 113–128.
[30] T. Rungrotmongkol, S. Hannongbua, A. Mulholland, Mechanistic study of HIV-1 reverse transcriptase at the active site based on QM/MM method, J. Theor. Comput. Chem. 3 (2004) 491–500.
[31] J. Florian, M.F. Goodman, A. Warshel, Computer simulation of the chemical catalysis of DNA polymerases: discriminating between

alternative nucleotide insertion mechanisms for T7 DNA polymerase, J. Am. Chem. Soc. 125 (2003) 8163–8177.
[32] R.C. Rittenhouse, W.K. Apostoluk, J.H. Miller, T.P. Straatsma, Characterization of the active site of DNA polymerase beta by molecular dynamics and quantum chemical calculation, Proteins: Struct. Funct. Genet. 53 (2003) 667–682.
[33] L. Yang, K. Arora, A.B. William, H.W. Samuel, T. Schlick, Critical role of magnesium ions in DNA polymerase beta's closing and active site assembly, J. Am. Chem. Soc. 126 (2004) 8441–8453.
[34] Y.G. Abashkin, J.W. Erickson, S.K. Burt, Quantum chemical investigation of enzymatic activity in DNA polymerase beta. A mechanistic study, J. Phys. Chem. B 105 (2001) 287–292.
[35] J.J.P. Stewart, MOPAC: a semiempirical molecular orbital program, J. Comput.: Aid. Mol. Des. 4 (1990) 1–105.
[36] H. Huang, R. Chopra, G.L. Verdine, S.C. Harrison, Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance, Science 282 (1998) 1669–1675.
[37] W. Saenger, Principles of Nucleic Acid Structure, Springer-Verlag, New York, 1983.
[38] D. Morikis, A.H. Elcock, P.A. Jennings, J.A. McCammon, Proton transfer dynamics of GART: the pH-dependent catalytic mechanism examined by electrostatic calculations, Protein Sci. 10 (2001) 2379–2392.
[39] D. Morikis, A.H. Elcock, P.A. Jennings, J.A. McCammon, The pH dependence of stability of the activation helix and the catalytic site of GART, Biophys. Chem. 105 (2003) 279–291.
[40] M.R. Sawaya, R. Prasad, S.H. Wilson, J. Kraut, H. Pelletier, Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism, Biochemistry 36 (1997) 11205–11215.
[41] N. Kaushik, V.N. Pandey, M.J. Modak, Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis: dynamics of the dNTP binding pocket, Biochemistry 35 (1996) 7256–7266.
[42] S.G. Sarafianos, V.N. Pandey, N. Kaushik, M.J. Modak, Glutamine 151 participates in the substrate dNTP binding function of HIV-1 reverse transcriptase, Biochemistry 34 (1995) 7207–7216.
[43] D. Harris, N. Kaushik, P.K. Pandey, P.N. Yadav, V.N. Pandey, Functional analysis of amino acid residues constituting the dNTP binding pocket of HIV-1 reverse transcriptase, J. Biol. Chem. 273 (1998) 33624–33634.
[44] C.L. Brooks, M. Karplus, Solvent effects on protein motion and protein effects on solvent motion: dynamics of the active site of lysozyme, J. Mol. Biol. 208 (1989) 159–181.
[45] J. Zurek, A.L. Bowman, W.A. Sokalski, A.J. Mulholland, MM and QM/MM modeling of threonyl-tRNA synthetase: model testing and simulations, Struct. Chem. 15 (2004) 405–414.
[46] C.L. Brooks, M. Karplus, Deformable stochastic boundaries in molecular dynamics, J. Chem. Phys. 79 (1983) 6312–6325.
[47] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.
[48] CHARMM. Available from the CHARMM Development Project H, University C, MA.
[49] A.D. MacKerell, 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, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[50] A.D. MacKerell, J. Wiorkiewicz-Kuczera, M. Karplus, An all-atom empirical energy function for the simulation of nucleic acids, J. Am. Chem. Soc. 117 (1995) 11946–11975.
[51] N. Foloppe, A.D. Mackerell, All-atom empirical force field for nucleic acids. I: Parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. 21 (2000) 86–104.
[52] A.D. Mackerell, N.K. Banavali, All-atom empirical force field for nucleic acids. II: Application to molecular dynamics simulations of DNA and RNA in solution, J. Comput. Chem. 21 (2000) 105–120.
[53] A.T. Brunger, M. Karplus, Polar hydrogen positions in proteins: empirical energy function placement and neutron diffraction comparison, Proteins 4 (1988) 148–156.
[54] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[55] K.E. Ranaghan, L. Ridder, B. Szefczyk, W.A. Sokalski, J.C. Hermann, A.J. Mulholland, Insights into enzyme catalysis from QM/MM modelling: transition state stabilization in chorismate mutase, Mol. Phys. 101 (2003) 2695–2714.
[56] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, J. Comput. Phys. 23 (1977) 327–341.
[57] W.F. van Gunsteren, H.J.C. Berendsen, Algorithms for macromolecular dynamics and constraint dynamics, Mol. Phys. 34 (1977) 1311–1327.
[58] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[59] M.J.S. Dewar, C. Jie, AM1 parameters for phosphorus, Theochem: J. Mol. Struct. 56 (1989) 1–13.
[60] J.J.P. Stewart, Optimization of parameters for semiempirical methods. I: Method, J. Comput. Chem. 10 (1989) 209–220.
[61] J.J.P. Stewart, 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 (1991) 320–341.
[62] M.C. Hutter, J.R. Reimers, N.S. Hush, Modeling the bacterial photosynthetic reaction center. 1: Magnesium parameters for the semiempirical AM1 method developed using a genetic algorithm, J. Phys. Chem. B 102 (1998) 8080–8090.
[63] M.J. Field, P.A. Bash, M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, J. Comput. Chem. 11 (1990) 700–733.
[64] A. Frank, P. Carloni, Ab initio molecular dynamics studies on HIV-1 reverse transcriptase triphosphate binding site: implications for nucleoside analog drug resistance, Protein Sci. 9 (2000) 2535–2546.
[65] D.M.Y. Xabier Lopez, Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus, Theor. Chem. Acc. 109 (2003) 149–159.