Theor Chem Account (2008) 120:479–489
DOI 10.1007/s00214-008-0440-9

REGULAR ARTICLE

Characterization of the active site of yeast RNA polymerase II by DFT and ReaxFF calculations

Rui Zhu · Florian Janetzko · Yue Zhang · Adri C. T. van Duin · William A. Goddard III · Dennis R. Salahub

Received: 31 October 2007 / Accepted: 14 March 2008 / Published online: 8 April 2008

Abstract Most known DNA-dependent RNA polymerases (RNAPs) share a universal heptapeptide, called the NADFDGD motif. The crystal structures of RNAPs indicate that in all cases this motif forms a loop with an embedded triad of aspartic acid residues. This conserved loop is the key part of the active site. Based on the crystal structures of the yeast RNAP II, we have studied this common active site for three cases: (1) single RNAP, (2) pre-translocation elongation complex, and (3) post-translocation elongation complex. Here we have applied two different modeling methods, the GGA density functional theory method (PBE) of quantum mechanics (QM) and the ReaxFF reactive force field. The QM calculations indicate that the loop shrinks from pre- to post-translocation and expands from post- to pre-translocation. In addition, PBE MD simulations in the gas phase at 310 K shows that the loop in the single-RNAP case is tightly connected to a catalytic Mg^{2+} ion and that there is an ordered hydrogen bond network in the loop. The corresponding ReaxFF MD simulation presents a less stable loop structure, suggesting that ReaxFF may underestimate the coordinating interactions between carbonyl oxygen and magnesium ion compared to the gas phase QM. However, with ReaxFF it was practical to study the dynamics for a much more detailed model for the post-translocational case, including the complete loop and solvent. This leads to a plausible reactant-side model that may explain the large difference in efficiency of NTP polymerization between RNA and DNA polymerases.

Keywords Yeast RNA polymerase II · NADFDGD motif · Nucleotidyl transfer · Molecular dynamics simulations · ReaxFF

Contribution to the Nino Russo Special Issue.

R. Zhu · F. Janetzko · Y. Zhang · D. R. Salahub
Department of Chemistry, University of Calgary,
Calgary, AB, Canada
e-mail: rzhu@ucalgary.ca

Y. Zhang
School of Chemistry and Materials Science,
Shaanxi Normal University, Xi'an, People's Republic of China

A. C. T. van Duin · W. A. Goddard III
Materials and Process Simulation Center,
California Institute of Technology, Pasadena, CA, USA

D. R. Salahub (✉)
Biological Sciences 556, Institute for Biocomplexity
and Informatics, University of Calgary,
2500 University Drive NW, Calgary, AB,
Canada, T2N 1N4
e-mail: dennis.salahub@ucalgary.ca

D. R. Salahub
Institute for Biocomplexity and Informatics, University of Calgary,
Calgary, AB, Canada

D. R. Salahub
Institute for Sustainable Energy, Environment and Economy,
University of Calgary, Calgary, AB, Canada

1 Introduction

DNA and RNA polymerases are both high-fidelity polynucleotide polymerases. DNA polymerases replicate the genome accurately for maintaining genomic stability in evolution. RNA polymerases catalyze the synthesis of messenger RNAs accurately based on the DNA template; otherwise the information contained in the DNA would be lost in gene expression. The catalytic mechanism proposed for both DNA and RNA polymerases involves the oxygen atom of the 3′-OH group of the RNA/primer acting as a nucleophile to form

a phosphodiester bond with the α phosphate of the matched (deoxy)nucleoside triphosphate (d-NTP/NTP) while the other two phosphates leave as a pyrophosphate. An illustration of this nucleotidyl transfer scheme in the context of RNA polymerases is presented in Fig. 1b.

From crystal structure analyses, these two classes of polymerases share a similar active site [1] with two or three conserved aspartic acid residues. On the other hand, RNA polymerases (RNAPs) also have some unique structural features at the active site. An extensive comparison between the amino acid sequences of the known largest subunit of the RNAPs of different organisms (including prokaryotes, eukaryotes, and cytoplasmic DNA viruses) indicates that all these polypeptides possess a universal heptapeptide: Asn Ala Asp Phe Asp Gly Asp, called the NADFDGD motif [2]. By comparing the crystal structures of the thermus aquaticus RNAP (PDB: 1HQM) [3], thermus thermophilus RNAP (PDB: 1IW7) [4], and yeast RNAP II (PDB: 1WCM) [5], we learn that, in all cases, the NADFDGD motif forms a loop with an embedded triad of aspartic acid residues [6] and that this triad always holds a Mg^{2+} ion, as illustrated in Fig. 2. This conserved triad is the key part of the active site of these RNAPs. The DNA polymerases, however, do not have this kind of common motif. They thus do not have such a loop-shaped part at the active site.

Recent observations of crystal structures of a series of yeast RNAP II complexes in different transcriptional phases [7,8] provide a good opportunity to study this common loop. The primary purpose of this work is to explore the structure of the yeast RNAP II active site in different phases at the atomic level, focusing on the loop structure. Specifically, we studied the active site both in a single RNAP and in an RNAP/DNA/transcript/NTP complex. For the latter case, there are two further situations: the complex after the nucleotidyl transfer (pre-translocation of RNAP as illustrated in Fig. 1a) and the complex before the transfer (post-translocation of RNAP as illustrated in Fig. 1b). A nucleotide residue is added to the transcript after each translocation of RNAP. Thus, we studied three cases in total. For both the single-RNAP and the pre-translocational cases, the three aspartate residues coordinate only one Mg^{2+} ion (Mg^{2+}(I) in Fig. 1). This metal ion appears to play the very important role of supporting the formation of a nucleophile in the catalysis. The post-translocational case shown in Fig. 1b corresponds to a reactant-side complex, where a second catalytic Mg^{2+}(II) ion has been proposed to come into the active site with the NTP. Mg^{2+}(II) and Mg^{2+}(I) appear to stabilize the penta-coordinated intermediate that forms during the nucleophilic attack, and Mg^{2+}(II) assists the leaving of the pyrophosphate. As a result, these two metal ions help to catalyze efficiently the nucleotidyl transfer in the NTP addition cycle, which is called the two-metal-ion mechanism [1,9].

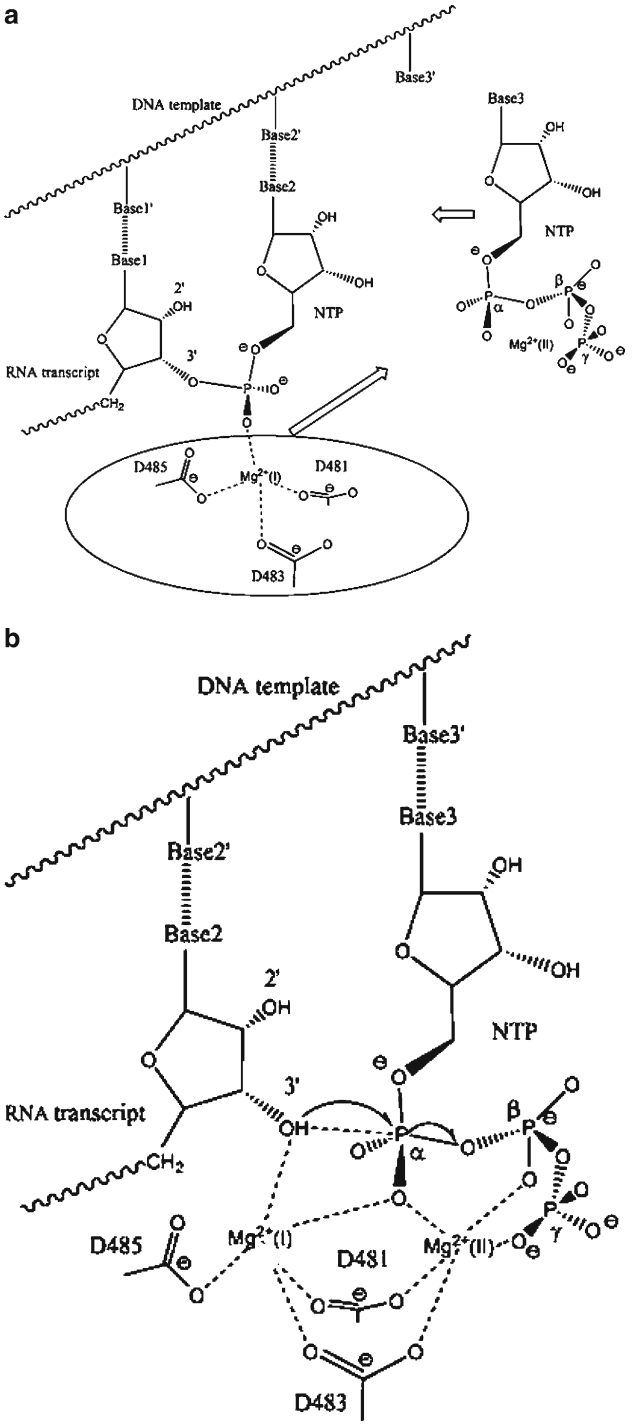

Fig. 1 The proposed NTP-addition scheme involved in RNA polymerases. **a** Pre-translocation complex: the *circled polymerase* part (including a permanent Mg^{2+}(I)) moves forward to the next base site, being coupled with the insertion of an NTP (holding another Mg^{2+}(II)). **b** Post-translocation with a matched NTP showing the mechanism of nucleotidyl transfer reaction (modified from Ref. [1].)

To investigate the loop structure at the three different cases mentioned above, we first performed molecular modeling at the quantum chemical level using the PBE density functional

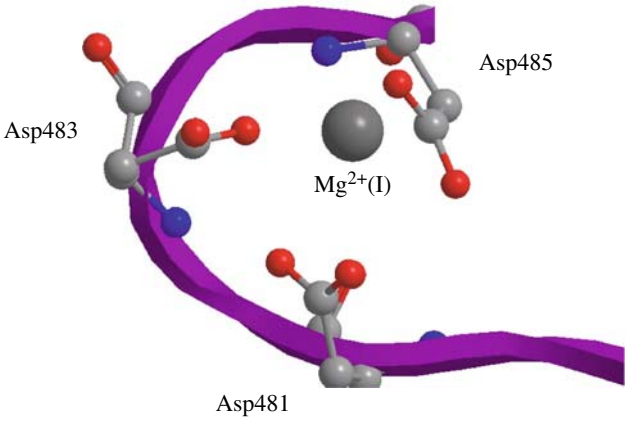

Fig. 2 Structure of the conserved NADFDGD motif of RNAPs which forms a common loop with a triad of aspartate residues holding a catalytic magnesium ion. This structure is obtained from the crystal structure (PDB:1I6H)

theory (DFT) method [10], which is expected to reveal the important internal interactions involved in the loop. Three models derived from the yeast RNAP II crystal structures were used. The models for the single-RNAP, pre-translocational, and post-translocational cases have 57, 91, and 108 atoms, respectively. We notice that even though the size of the pre-translocational complex model is too large for the PBE method to handle, it is still too small to capture some important interactions involved in the catalysis. Therefore, we also constructed a more detailed model of 250 atoms for the post-translocational case. We then studied this model system in a 30 × 30 × 30 Å cell randomly filled with 400 water molecules by using the recently developed reactive force field (ReaxFF). ReaxFF is a first principles-based reactive force field [11,12], in which the parameters are fitted to a quantum mechanics training set. The main purpose of using ReaxFF is to provide a reasonable model for our study of possible reaction pathways involved in the nucleotide-addition catalytic cycle. We also intend to derive some simplified models from this large ReaxFF model for the catalytic reaction modeling at the DFT level.

2 Models and methods

2.1 DFT-level modeling

After observing several yeast RNAP II complex crystal structures, we found that only five residues of the NADFDGD motif, i.e., Asp Phe Asp Gly Asp, are tightly involved in the active site. Thus, we focused only on this amino acid sequence, called the DFDGD motif, in the DFT-level modeling. We first built a single yeast RNAP II active site model, called model 1, from the pre-translocation crystal structure (PDB: 1I6H, resolution 3.3 Å). See the next paragraph for an

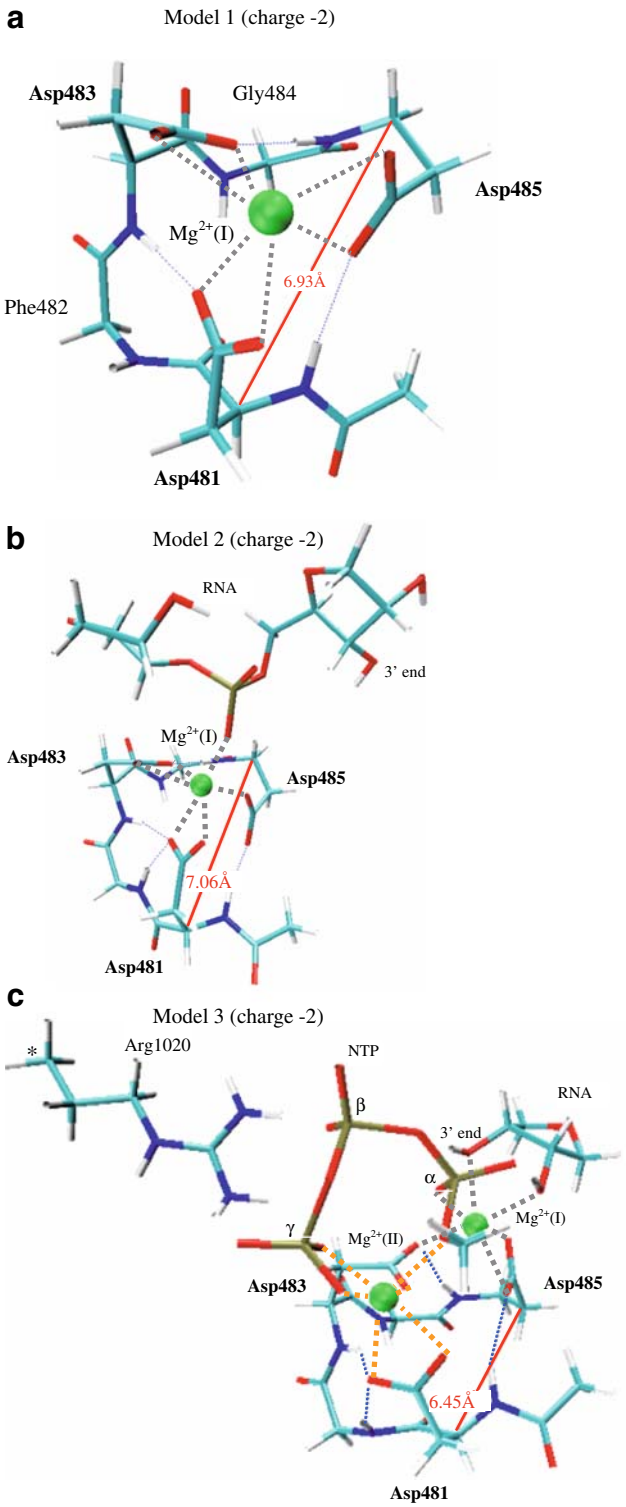

Fig. 3 PBE-level optimized structures of **a** model 1, **b** model 2, and **c** model 3 for the single RNAP, pre-translocation and post-translocation cases, respectively. Each model system has a net charge. The carbon atom marked with ∗ is fixed in the optimization. *Slim dashed lines* denote hydrogen bonds formed in the loop; *thick dashed lines* denote coordination bonds; *solid lines* denote the opening of the loop. Note that Arg1020 belongs to subunit β and other residues to subunit α. Please view the electronic version for a better illustration in color

explanation of using the pre-translocation structure to build a model in a single RNAP. In this model, as illustrated in Fig. 3a, the DFDGD motif is further simplified to a sequence of Asp Gly Asp Gly Asp. We consider only the full aspartic acid residues and peptide connections between them, but ignore the effect of the phenyl group of the phenylalanine residue. Note that our model is different from those active site models of DNA polymerases, where the aspartate residues were simplified to formate groups [13,14] or only the carboxylates of the aspartate residues were involved in the quantum chemical calculations [15], and where no peptide bonds were involved. It is important to note that both the size of the model and the theoretical methodology can influence the final result. DFT calculations provide a detailed description of the electronic structure of the model system. However, it is often necessary to omit many structural features to decrease the calculation costs. This might make the DFT description lead to results that do not apply to the real system.

As illustrated in Fig. 3b, model 2, representing a pre-translocation complex, is model 1 combined with a simplified model of the RNA transcript taken from the crystal structure 1I6H. It is worth noting that we used the same pre-translocation elongation complex (1I6H) to build the initial structures of models 1 and 2. The only difference between the two models is that model 1 does not include the RNA segment as in model 2. The comparison between the two optimized geometries allows us to investigate the changes of the loop due to the addition of the RNA transcript to it. Model 3 in Fig. 3c, depicting a post-translocation complex, was derived from the post-translocation elongation complex (PDB: 1R9S, resolution 4.3 Å). It contains the model 1 structure, a second Mg^{2+}(II) ion, a simplified RNA, a simplified arginine residue, and a simplified NTP substrate. It should be noticed that this post-translocation crystal structure was actually obtained by combining the substrate data in post-translocation with the active site motif data in pre-translocation (1I6H). In other words, the initial geometry of model 3 has exactly the same loop geometry as in the initial geometries of models 1 and 2. This allows us to investigate the changes of the loop structure before and after translocation of RNAP.

For the modeling part, the geometry optimization for all the three models was first performed using the linear combination of Gaussian-type orbital (LCGTO) Kohn–Sham (KS) DFT program deMon2k [16]. The GGA PBE96-PBE exchange-correlation functionals were employed [10]. The double-zeta valence polarization (DZVP) [17] basis set was used in combination with the GEN-A2* auxiliary function set [18] for the variational fitting of the Coulomb potential. The structure optimizations were done using a Quasi-Newton method with analytical gradients. Next, we performed for model 1 an MD simulation in the gas phase at the same DFT level for a simulation time of about 5 ps starting from the optimized structure. A Nose–Hoover thermostat was used and the simulations were performed with time steps of 0.5 fs. The equilibrium temperature was set to 310 K.

Table 1 ReaxFF bond parameters for Mg–O interactions. For a full description of these parameters and the ReaxFF functional form see Ref. [12]

Interaction	De_{sigma}	p_{be1}	p_{ovun1}	p_{be2}	p_{bo1}	p_{bo2}
Mg–O	87.0227	0.0030	0.0250	0.0087	−0.0439	6.6073

Table 2 ReaxFF off-diagonal van der Waals and bond parameters for Mg-O interactions

Interaction	Dij	R_{vdW}	Alfa	r_{sigma}
Mg–O	0.0809	1.700	11.4606	1.5177

2.2 ReaxFF-level modeling

To build a more detailed post-translocation model in the yeast RNA polymerase II, we derived a backbone model of 250 atoms, called model 4, from a recently published post-translocation crystal structure (PDB: 2NVZ, resolution 4.3 Å). This crystal structure is a refined version of 1R9S, containing more accurate information about residues at the active site. The intention of building model 4 is to get a reasonable model for the investigation of the catalytic nucleotidyl transfer mechanism. The initial geometry of model 4, as illustrated in Fig. 8a, features the loop structure of the complete NADFDGD motif, two Mg^{2+} ions, two arginine residues, one histidine residue, and the whole incoming NTP at the addition site. The net charge of model 4 is zero. The backbone system was first energy minimized with the ReaxFF program. Next, the minimized structure was equilibrated at 310 K in a 30 × 30 × 30 Å cell containing 400 water molecules with the ReaxFF program. Some atoms as indicated by arrows in Fig. 8a were fixed in the MD simulation. A Berendsen thermostat with a temperature damping constant of 100 fs was used to control the system temperature. A MD-time step of 0.25 fs was used in the MD simulation. ReaxFF bond parameters for Mg–O interactions are given in Tables 1, 2 and 3.

3 Results and discussion

3.1 Optimized geometries at the DFT level

The fully optimized geometry of model 1 at the DFT level is shown in Fig. 3a. The three aspartate residues are numbered as in the yeast RNAP II. As expected, the Mg^{2+}(I) ion is held by the three aspartate residues through coordination with

Table 3 ReaxFF angle parameters for Mg/O/H interactions.

Interaction	$\text{Theta}_{o,o}$	p_{val1}	p_{val2}	p_{val4}	p_{val7}
O–Mg–O	0[a]	9.232	0.1	1.092	1.0
Mg–O–Mg	0	25	8.0	3.0	1.0
H–O–Mg	66.04	5.0	1.0	1.25	1.0
H–Mg–O	0	0.5	0.1	3.0	1.0
O–O–Mg	70.0	20.0	1.0	1.25	1.0

[a] leads to an equilibrium angle of 180–$\text{Theta}_{o,o}$

Table 4 Distance measurements (in Å) in models 1–3

	Asp485	Asp481	Asp483
Model 1			
Hydrogen bond[a]	2.05	1.93	1.88
		–	
Mg_1–O	2.11 2.19	2.13 2.20	2.11 2.21
Model 2			
Hydrogen bond	1.85	2.14	1.99
		1.63[b]	
Mg_1–O	2.00[c] 2.02	2.21 2.22	2.21 2.28
Model 3			
Hydrogen bond	2.30	2.47	2.71
		1.77	
Mg_1–O	2.14 2.22	–	2.06
	P_αO:2.10, 2-OH:2.21, 3-OH:2.28		
Mg_2–O	2.13	2.13 2.28	2.10
	P_αO: 2.25, P_γO:2.07		

Geometries are optimized at the PBE level

[a] For example, the hydrogen bond for Asp481 denotes that one formed by one of the carboxylate oxygen atoms of Asp481 and the N-hydrogen of Asp483

[b] The second hydrogen bond for Asp481 is formed by the same carboxylate oxygen of Asp481 and the N-hydrogen of Phe482

[c] This value is for the distance between Mg_1 and one of the non-bridging oxygen atoms of the phosphate

three carboxylates, forming a favorable six-ligand geometry. Interestingly, an ordered hydrogen-bond network is also clearly shown in the structure. One of the two carboxylate oxygen atoms of each Asp is hydrogen-bonded to the backbone amide group of a nearby Asp: Asp485–O–Asp481–H, Asp481–O–Asp483–H, Asp483–O–Asp481–H. The distances between the Mg and O atoms and the lengths of the hydrogen bonds are given in Table 4. This energy-favored structure implies that the DFDGD motif has an intrinsic ability to form an ordered triad structure through interactions with an Mg^{2+} ion.

As shown in Fig. 3b, the optimized model 2, modeling the pre-translocation complex, has an RNA part consisting of two sugars and a scissile phosphate. Note again that we used the same pre-translocation elongation complex (1I6H) to build models 1 and 2, i.e., model 2 = model 1 + the RNA part. Thus, the comparison between the two optimized geometries allows us to see differences between the pre-translocational and the single RNAP cases. Compared with model 1, the non-bridging oxygen of the phosphate in model 2, replacing one carbonyl oxygen of Asp485, coordinates the Mg^{2+}(I) ion. A quite similar situation is shown in the corresponding pre-translocation crystal structure 1I6H, where it is, however, a bridging oxygen of the phosphate that coordinates the Mg^{2+}(I) ion. As a result of this change in the coordination of the magnesium, the Asp485–O–Asp481–H hydrogen bond gets stronger in the pre-translocational phase than in the single RNAP case. We can see this change from the fact that the length of the hydrogen bond decreases from 2.05 (model 1) to 1.85 Å (model 2). In addition, a strong fourth hydrogen bond is formed in model 2 by the carboxylate oxygen of Asp481 and the N-hydrogen of Phe482. The distances between the Mg and O atoms and the lengths of the hydrogen bonds in model 2 are also listed in Table 4. Also worth noting is that the loop opening increases a little from 6.9 (model 1) to 7.1 Å (model 2) due to the change of the coordinating state caused by the RNA part. The loop opening is denoted by a solid line in Fig. 3.

By fixing one carbon atom of Arg1020, we obtained an optimized geometry of model 3 (Fig. 3c) for the post-translocational case. Arg1020 belongs to a different subunit (β) which is assumed to be robust in the studied case. Such a restriction allows free rotation of the functional group of Arg1020 without moving Arg1020 away from the subunit β. The substrate part obtained is similar to the corresponding crystal structure in the post-translocation complex (1R9S). Also, the distance between the two magnesium ions is 4.32 Å, which is in qualitative agreement with the crystal data, 4.15 Å. As mentioned earlier in Sect. 2, the initial geometries of models 3 and 2 have exactly the same loop structure. Interestingly, the optimized geometry of model 3 shows that the motif loop appears to shrink compared with the optimized model 2. As indicated in Fig. 3b, c, the loop opening decreases from 7.1 to 6.5 Å. This indicates a shrinking of the loop from the pre- to post-translocation and an expanding from the post- to pre-translocation. We also obtained another optimized geometry of model 3 which is about 25 kcal/mol higher in energy than that shown in Fig. 3c when using a slightly different initial conformation. In this geometry, the distance between the two magnesium ions is only 3.70 Å, much shorter than the crystal value. Similar to the one illustrated in Fig. 3c, however, a small loop was also obtained. We suppose that this shrinking movement may play an important role in the catalysis. As the Mg^{2+}(II) ion moves over the loop with the NTP in the post-translocational phase, the shrinking of the loop lifts up the Mg^{2+}(I) ion close to the 3′-OH group and the α phosphate of the NTP, facilitating the nucleophilic attack. Note that we cannot get this loop change by comparing 1I6H and 1R9S since they have the same loop

geometry. However, we do see this shrinking tendency by comparing 1I6H and 2NVZ. Note that 2NVZ is the refined version of 1R9S. The loop opening changes from 6.8 (1I6H) to 6.2 Å (2NVZ). In addition, the distances between the Mg and O atoms and the lengths of the possible hydrogen bonds in our model 3 are presented in Table 4. It is found that only one hydrogen bond is clearly present in this case. The Mg^{2+} ion prefers six ligands arranged in an octahedral geometry with 2.05–2.25 Å coordination distances [19]. A structural feature of model 3 is that both Mg^{2+}(I) and Mg^{2+}(II) ions form the most favorable six-ligand configuration. We can see from Table 4 that the coordination distances are roughly in the range of 2.05–2.25 Å.

The above results imply that the motif loop of RNAP appears not to be fixed during the transcription process. Specifically, it shrinks when the RNAP is absorbing an NTP and expands when the RNAP is completing the addition of a nucleotide. This rhythmic movement may adjust the orientation of the Mg^{2+}(I) ion in the catalytic process, facilitating the nucleophilic attack. In addition, we note that the model systems studied here are all negatively charged. The net charges of models 1–3 are −1, −2, and −2, respectively. This is due to the three conserved aspartate residues (−3 charge) of the active site and the triphosphate (−4 charge) of the incoming NTP. The former is tightly connected to the Mg^{2+}(I) ion in the active site and the latter effectively brings the Mg^{2+}(II) ion into the active site. However, the extra negative potential makes the system unstable. Thus, those hydrogen bonds in models 1–3 seem to adjust the interactions of the negative oxygen atoms and the positive Mg^{2+} ion(s).

While the three models used above are quite simple, studying them is a reasonable first step toward understanding the features of the common loop. We ignored important interactions such as the solvation effect of water molecules and the polarization effects of the surrounding enzyme. These interactions may function significantly in the catalytic process. Though model 3 has 108 atoms, it is still not accurate enough to study the catalytic mechanism. For example, model 3 in Fig. 3c suggests the coordination of both 2′-OH and 3′-OH groups of RNA by Mg^{2+}(I) ion, which is not a good structure for the 3′-OH nucleophilic attack. Thus, we need to build a more detailed model with more atoms for the post-translocational case. In the next section, we build a model which includes more residues at the active site, a more detailed RNA, and a complete substrate. Furthermore, this backbone was studied in a cell filled with water molecules. Due to the large size of this model system, we turned to ReaxFF MD simulations.

DFT calculations provide an accurate description of the electronic structure of the model system so that DFT can be expected to provide a much better description of the reaction mechanism for the model system. On the other hand, to make the calculations practical for DFT it is necessary to omit many structural features from the model. The advantages of using a force field are that we can include the full protein and the solvent and ions and average the structures over the dynamics at the appropriate temperature. The special advantage of ReaxFF is that this averaging can also be done as the reaction is proceeding.

3.2 MD simulations

There are two parts in this section. In the first part, we performed a PBE MD simulation and also a ReaxFF MD simulation both in the gas phase for model 1. One purpose is to see how robust model 1 is in the gas phase at 310 K. Note that the model 1 structure shown in Fig. 3a is just a minimum at 0 K. The other purpose is to test the performance of ReaxFF by comparing the ReaxFF results with the PBE ones. In the second part, we carry out a ReaxFF MD simulation for a more detailed model of the post-translocation complex in a water environment.

3.2.1 The PBE and ReaxFF MD simulations for model 1

Both the PBE and ReaxFF MD simulations ran about 5 ps. The final temperature was set to 310 K. The trajectory of the average temperature and the potential energy of model 1 are shown in Fig. 4 for the two simulations. In the first comparison between the two simulations, we chose the strong interactions, i.e., the covalent bonds between atoms C, N, O, and H. Six covalent bonds as labeled in Fig. 5 were compared in the two simulations. The statistics of the bond lengths and the corresponding crystal structure data from 1I6H are given in Table 5. First, we can see that the average bond lengths obtained from the PBE MD simulation are in excellent agreement with the corresponding crystal data. Second, the relative fluctuations of the bond lengths in the two simulations all match well except for the C–H and the C–C bonds. Third, the absolute fluctuations of the bond lengths in the ReaxFF simulation are slightly bigger than those in the PBE simulation, especially for the O–C and C–N bonds. These comparisons indicate that the PBE method is suitable for describing these strong interactions and that the descriptions of these strong interactions are similar with ReaxFF and PBE.

For the PBE geometry optimization, the calculated geometry is just a minimum energy structure. The observed X-ray structure is averaged over the time scale of the experiment and might average over more than one minimum. Thus, the equilibrium geometry from an MD simulation at the appropriate temperature should be the appropriate structure to compare to the X-ray structure.

Next, we compare the Mg–O coordination interactions in the two simulations. All the six coordination bonds in model 1 were compared. The fluctuations of the Mg–O distances are plotted in Fig. 6. As can be seen, the fluctuations

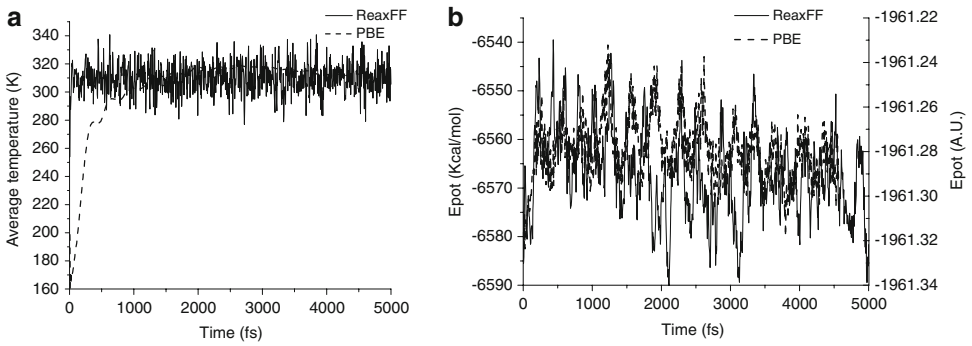

Fig. 4 Evolution of **a** the temperature and **b** the potential energy of the system in dependence of the simulated time in the molecular dynamic simulations of model 1 using the ReaxFF and deMon2k PBE methods. The temperature at each point of the trajectory for the ReaxFF simulation is actual temperature at that time point, while that for the PBE calculation is averaged over the simulated time passed. The current potential energy is given for each time point (ReaxFF energies in Kcal/mol and PBE energies in A.U.)

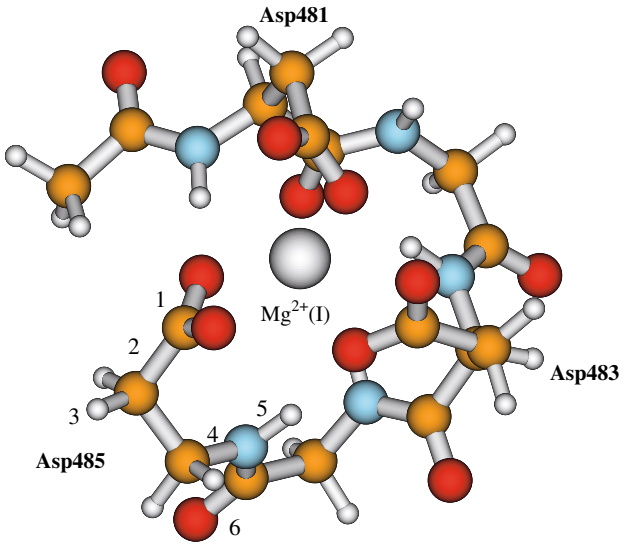

Fig. 5 Labels of selected bonds in model 1. The maximum and minimum lengths of these bonds during the ReaxFF and deMon2k PBE MD simulations are shown in Table 5 in order to compare the description of covalent bonds in both methods with each other

in all the six cases range from 1.9 to 2.6 Å in the PBE MD simulation, indicating that three bidentate carboxylates bind to Mg almost all the time. The interactions between each carboxylate O and the Mg^{2+} ion appear to be very strong, making this active site model very robust in the gas phase even at 310 K. On the ReaxFF side, however, the fluctuations of the Mg–O distances are much larger than those in the PBE simulation. As can be seen in Fig. 6, in the ReaxFF simulation, only one bidentate carboxylate binding remains (labels 1, 2), and the other two are replaced by monodentate binding (labels 3, 4 and labels 5, 6) after some MD steps. Our MD simulations show that the loop structure is more robust using the PBE method than using the ReaxFF force field. In principle, this breaking of the coordination could also occur in the PBE simulations if it were practical to carry out longer simulations. Such PBE MD simulations are extremely time-consuming. From comparisons of the PBE and ReaxFF MD trajectories, we see that the current implementation of ReaxFF seems to underestimate the coordination interactions. Thus, fluctuations of Mg^{2+} positions in the ReaxFF simulation are much larger than those in the PBE simulation.

Finally, to further investigate the loop, we studied the weak interactions involved in the loop, i.e., the hydrogen bonds, in the two MD simulations. The lengths of possible hydrogen bonds and the loop opening from the PBE simulation are plotted in Fig. 7. The four hydrogen bonds (labeled 2–5 in Fig. 7) which were observed in the optimized models 1–3 in

Table 5 Distance statistics (in Å) in model 1 from the MD simulations together with the corresponding crystal structure data

Label	Atoms	PBE			ReaxFF			Crystal structure data from 1I6H
		min∼max	max – min	mean	min∼max,	max – min,	mean	
1	O–C	1.23∼1.37	0.14	1.29	1.31∼1.44	0.13	1.36	1.25
2	C–C	1.42∼1.62	0.20	1.52	1.48∼1.61	0.13	1.53	1.53
3	C–H	1.02∼1.23	0.21	1.12	1.08∼1.18	0.10	1.13	n/a
4	C–N	1.37∼1.56	0.19	1.46	1.45∼1.63	0.18	1.52	1.46
5	N–H	0.96∼1.12	0.16	1.04	1.00∼1.14	0.14	1.07	n/a
6	C–O	1.21∼1.29	0.08	1.25	1.27∼1.34	0.07	1.30	1.23

Note that the "min∼max" and the "max – min" denote the absolute and relative fluctuations of a bond length, respectively

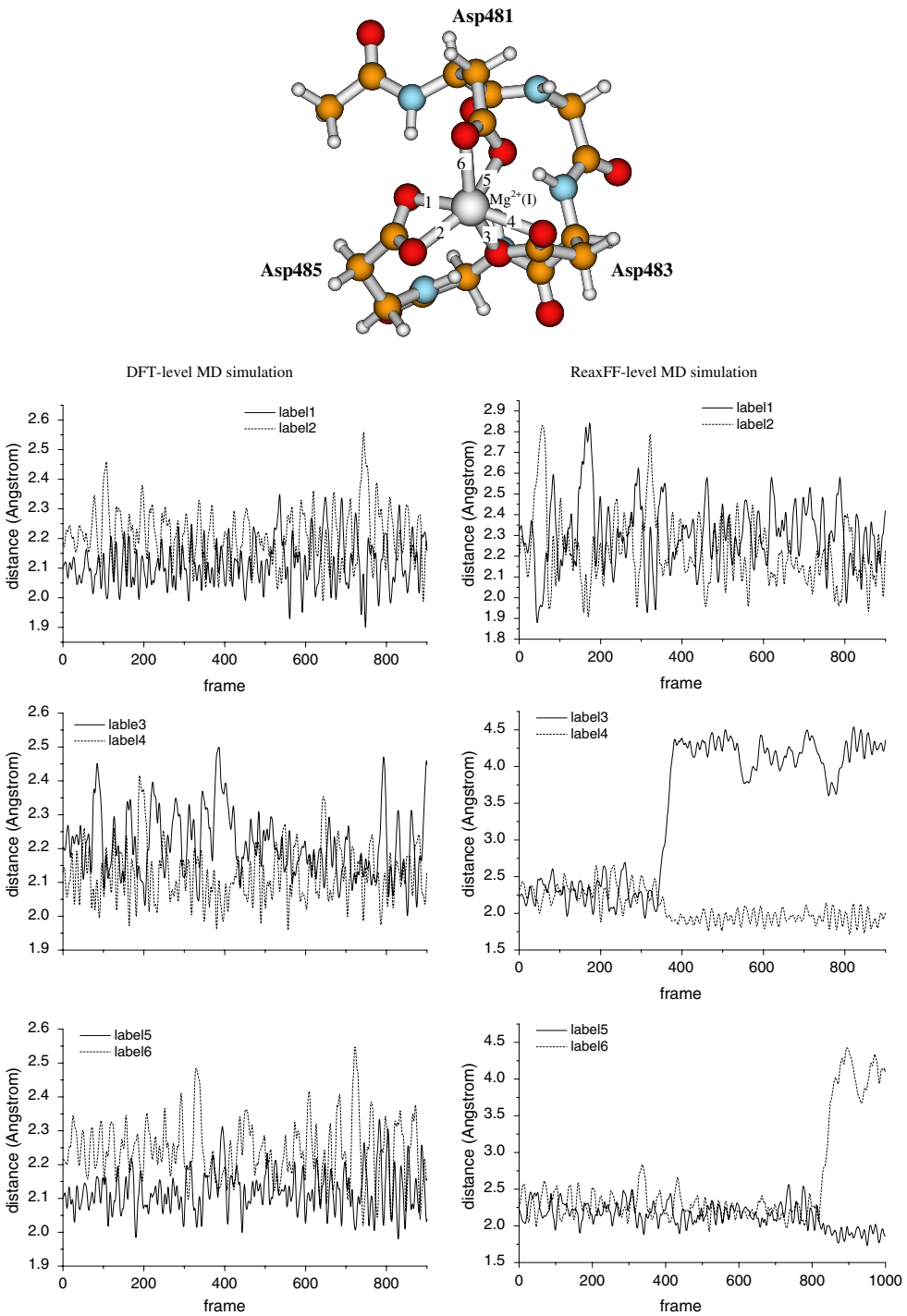

Fig. 6 Fluctuations of the six Mg–O distances (labels 1–6) in dependence of the simulated time during the MD simulations with deMon2k PBE (*left part*) and ReaxFF (*right part*) methods. One frame corresponds to a simulated time of 5 fs

Fig. 3 all show up in the PBE simulation. As seen in Fig. 7, the label 3 hydrogen bond seems to be the strongest one among them, which remains in the whole simulation. The weakest one seems to be the label 2 hydrogen bond, which interestingly is associated with the label 1 loop opening. Even more interestingly, the label 4 and label 5 hydrogen bonds are able to switch in the MD simulation. On the ReaxFF side, only two hydrogen bonds (labels 4 and 5) seem to remain, and the others were not present at all in the simulation (results not shown). In addition, the ReaxFF loop opening fluctuates much more than the PBE one. The former ranges from 5.0 to 8.7 Å, while the latter ranges from 6.7 to 8.3 Å. We suppose that these differences are closely related with the weak Mg–O interactions in ReaxFF.

In summary, comparing the PBE and ReaxFF MD simulations, we learn that ReaxFF works well for covalent

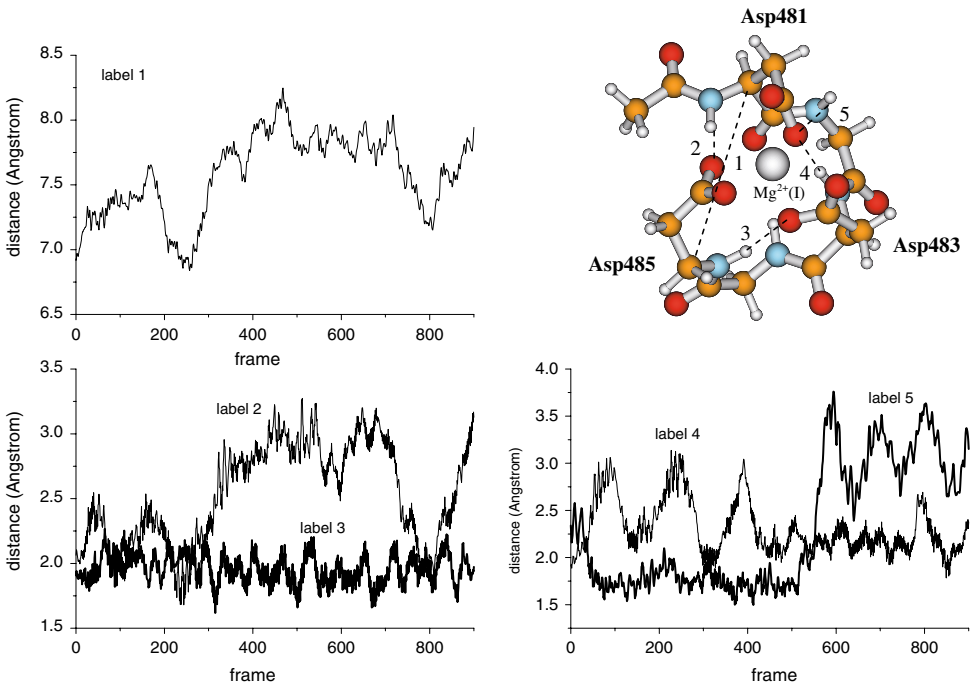

Fig. 7 Trajectories for a characteristic distance (label 1) describing the loop opening and for the four hydrogen bonds (labels 2–5) obtained from deMon2k PBE molecular dynamic simulations which show how the loop size and the bond lengths change during the MD run. Note that the hydrogen bonds label 4 and label 5 interchange after roughly 275 fs (550 frames). One frame corresponds to a simulated time of 5 fs

interactions between C, N, O, H, but underestimates the Mg–O coordination interactions compared with the PBE method. The PBE simulation indicates that the simple model 1 itself has a strong potential to form a robust loop in the gas phase at 310 K. However, we could not get such a robust loop in the corresponding ReaxFF simulation. From the strong interactions of covalent bonds to the intermediate interactions of coordination bonds to the weak interactions of hydrogen bonds, the differences between the two simulations strongly imply that the structural stability of the loop mainly comes from the Mg–O coordination interactions. These intermediate interactions then cause a good environment for the formation of the hydrogen-bond network in the loop.

3.2.2 The ReaxFF MD simulation for model 4

While the current version of ReaxFF needs to be further developed for a better description of the Mg–O interactions, we can still use this tool to do some helpful modeling work which may guide the reaction mechanism studies at the DFT level. As mentioned above, model 3 is not accurate enough to examine the catalysis. We suppose that some water molecules and other residues may significantly be involved in the nucleotide addition cycle. Unfortunately, all the yeast RNAP II crystal structures lack the information on water molecules close to the active site. How can we get this water and residue information? First, we need a more detailed crystal structure in post-translocation and we need to build a backbone model with more residues based upon it. Second, we put the backbone in a cell filled with water molecules and relax the system at 310 K, obtaining some equilibrium geometries. Finally, we build models which are used for DFT-level studies based on reasonable equilibrium geometries. In step 1, we can use the refined crystal structure 2NVZ to build a more detailed pre-translocation model. For step 2, the ReaxFF program can be used to perform the necessary simulations. In addition, before performing the last step, we can also use the ReaxFF program to directly examine some possible reaction pathways based on a reasonable equilibrium geometry. The obtained results could be very useful to further guide the corresponding DFT-level studies. While traditional force fields can handle more atoms than ReaxFF in the case studied, they cannot provide information on chemical bond changes like bond formation and bond cleavage. This is the main reason that we are using ReaxFF rather than traditional force fields in this work. Here, we only perform the first two steps.

The initial structure of model 4 derived from 2NVZ is given in Fig. 8a. A recent experimental study [8] stressed that a histidine residue located in a trigger loop, which comes close to a matched NTP as the NTP enters the active site, may literally trigger phosphodiester bond formation. We thus included this histidine residue 1085 into model 4. Another nearby arginine residue 766 close to the γ phosphate was also included. Besides the two additional residues, we used in model 4 a complete NTP substitute, the complete NAD-FDGD motif structure, and a base–sugar primer. Using the method mentioned in Sect. 2 we obtained several equilibrium geometries of model 4 at 310 K in a water environment.

One reasonable equilibrium geometry is illustrated in Fig. 8b. There are some obvious differences between this

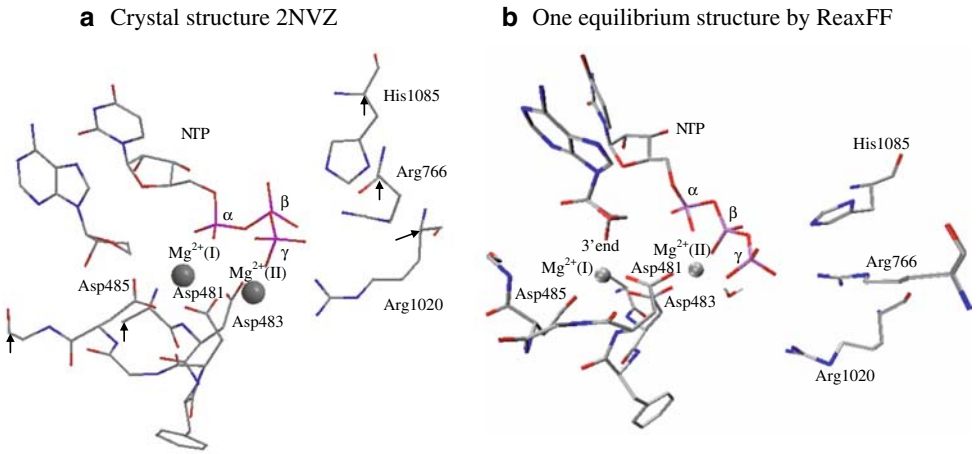

Fig. 8 **a** A reactant-side model (post-translocation) derived from the PDB:2NZV crystal structure; **b** One equilibrium structure observed in the corresponding ReaxFF MD simulation. Arg1020 and Arg766 belong to subunit β and other residues to subunit α. A close water molecule below Mg^{2+}(II) is also shown in **b**. Note that the 3′-OH group is missing in the crystal structure. This group is manually added to the model, but is not shown in **a** to stress the original crystal structure. Atoms indicated by *arrows* are fixed in the simulation

geometry and the corresponding crystal structure. For example, the two Mg^{2+} ions in the ReaxFF geometry move towards the RNA part in the equilibrium geometry. We suppose that this may be partly caused by the 3′-OH group added into model 4 which is missing in the crystal structure. We also note that the two Asps (481 and 483) in the crystal structure do not symmetrically coordinate the two Mg^{2+} ion, which may be due to the missing 3′-OH group as well. To test this assumption, we modified in the beginning of the simulation the positions of the two Asp residues so that they symmetrically coordinate the two Mg^{2+} ions. Interestingly, this conformation remained throughout the simulation.

The equilibrium structure obtained in Fig. 8b is quite similar to the corresponding X-ray structures observed for the analogous functional states in DNA polymerases [20,21]. Remarkably, not only the bridging configuration of the two Mg^{2+} ions and connecting carboxylates of two Asps is observed in the equilibrium structure, but also the arrangement of the triphosphate chain of the NTP substrate improves, so that the oxygen of the β phosphate becomes coordinated by Mg^{2+}(II). These features are the same as those observed in DNA polymerase structures.

The arrangement of the triphosphate chain could be a very important process because the coordination of the β phosphate is a crucial requirement for the nucleotidyl transfer according to the commonly accepted two-metal-ion mechanism. This may also explain the large difference in efficiency of NTP polymerization between RNA and DNA polymerases. The regular rate of nucleotide addition by RNA polymerases is about 10–50 nucleotides per second, while that for DNA polymerase is 500–1,000. Thus, it could be proposed that in DNA polymerases the active center is initially tuned for catalysis, while in RNA polymerases the active center needs some adjustment (as observed in the equilibrium structure, Fig. 8b) to be catalytically competent.

In addition, the three residues around the NTP provide some reasonable connections with the NTP. The histidine residue 1085 makes a hydrogen bond with the β phosphate; the arginine residue 766 links to the γ phosphate through a hydrogen bond; the arginine residue 1020 links to the γ phosphate through a water molecule. These positively charged residues may facilitate the phosphodiester bond formation and also the pyrophosphate group formation. A water molecule coordinating the Mg^{2+}(II) ion is also found in the simulation. This water molecule may further stabilize the penta-coordinated intermediate by forming the favorable six-ligand configuration. Interestingly, a water molecule coordinated by Mg^{2+}(II) in the *Thermus thermophilus* RNA polymerase elongation complex was observed in a better resolved X-ray structure (PDB: 2o5j) [22]. Therefore, we obtained a reasonable reactant-side model which could be used for the study of the catalytic mechanism. In a slightly modified model, we have indeed examined some reaction pathways in another work using the ReaxFF program. We are now studying some small models at the DFT level to re-examine the pathways.

4 Summary and conclusions

We studied the active site of the yeast RNAP II using both DFT and ReaxFF modeling methods. DFT calculations provide a description at a higher level of theory compared with the ReaxFF force-field calculations because the electronic structure of the system is explicitly taken into account in the DFT simulations. However, they are much more time

consuming than the ReaxFF simulations. We focused on the NADFDGD motif of the active site which forms a conserved loop in most known DNA-dependent RNAPs.

We first investigated a simplified version of the NADFDGD motif at the DFT-level in three cases: single-RNAP, pre-translocation, and post-translocation. The optimized geometries for the pre- and post-translocation cases suggest that the motif loop has a repeating movement during the transcription process. It shrinks when the RNAP is absorbing an NTP; it expands when the RNAP is completing the addition of a nucleotide. The same change of the loop opening between pre- and post-translocation is also shown in the corresponding crystal structures. We suppose that this rhythmic movement may adjust the orientation of the first catalytic Mg^{2+} ion in the catalytic process, facilitating the 3′-OH nucleophilic attack to the α phosphate of NTP.

In addition, we performed a DFT-level MD simulation for the single RNAP active site model. The results show that the loop together with an Mg^{2+} ion forms a robust structure in the gas phase at 310 K. The structural stability mainly comes from the Mg–O coordination interactions. The loop features a triad of aspartate residues which provide six carboxyl oxygen atoms possibly coordinating the Mg^{2+} ion. A hydrogen-bond network in the loop is clearly shown in the MD simulation, too. Those hydrogen bonds form to adjust the interactions of the negative carboxyl oxygen atoms and the positive Mg^{2+} ion and stabilize the system. We then performed a similar MD simulation using the ReaxFF program. The obtained loop structure is not as robust as the one shown in the DFT MD simulation, indicating that the current version of ReaxFF underestimates the Mg–O coordination interactions. It is worth mentioning that the ReaxFF Mg-parameters were derived from MgO–condensed phase properties; we should refine these parameters to better describe the competition between carboxylate, hydroxyl, and water binding to Mg^{2+}.

Our motivation for studying the RNAP active site is to examine the catalytic mechanism of nucleotidyl transfer, determining the enzyme basis of the features of transcriptional NTP polymerization. The DFT-level model for the post-translocational case is a reactant-side model (108 atoms) which we used to explore reaction pathways. However, we found that this model is too simplified for this task. What we need is a model providing a detailed description of the full protein, while also including the effects of water molecules in the reaction. This requires over 50,000 atoms in the model, far too many for DFT methods. Thus, we proposed a strategy to overcome these difficulties: starting from a more detailed active-site model (250 atoms), we use ReaxFF MD simulations to obtain equilibrium structures for a cell filled with water molecules. Starting with this reasonable equilibrium structure, we then use DFT to build a small model. In addition, we used ReaxFF directly to efficiently study several reaction pathways. This allows us to gain new insights into the mechanism. In addition, it leads to results useful for guiding subsequent DFT-level modeling studies.

Acknowledgements We thank an anonymous reviewer for valuable suggestions. In particular, the proposed rationale for the relative rates of DNA and RNA polymerization was stimulated by the reviewer's comments. We would like thank Dr. Hong Guo and Dr. Zhitao Xu for valuable discussions. We also thank WestGrid for assistance with calculations and the Natural Sciences and Engineering Research Council of Canada for support of this work 10174. F. J. thanks the Deutsche Forschungsgemeinschaft (DFG) for a fellowship (Forschungsstipendium). Partial support was also provided by NSF (CTS-0608889) and NIH (RO1 CA 112293-01). We wish Professor Nino Russo many happy returns on the occasion of his 60th birthday. D. R. S. has enjoyed many years of scientific collaboration and camaraderie with Nino and hopes there are many more such years to come.

References

1. Steitz TA (1998) Nature 391:231–232
2. Sonntag K-C, Darai G (1996) Virus Genes 11:271–284
3. Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, Severinov K (2001) Proc Natl Acad Sci USA 98:892–897
4. Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S, Yokoyama S (2002) Nature 417:712–719
5. Armache K-J, Mitterweger S, Meinhart A, Cramer P (2005) J Biol Chem 280:7131–7134
6. Sosunov V, Zorov S, Sosunova E, Nikolaev A, Zakeyeva I, Bass I, Goldfarb A, Nikiforov V, Severinov K, Mustaev A (2005) Nucleic Acids Res 33:4202–4211
7. Westover KD, Bushnell DA, Kornberg RD (2004) Cell 119:481–489
8. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006) Cell 127:941–954
9. Steitz TA (1993) Curr Opin Struct Biol 3:31–38
10. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865–3868
11. van Duin ACT, Dasgupta S, Lorant F, Goddard WAIII (2001) J Phys Chem A 105:9396–9409
12. Nielson KD, van Duin ACT, Oxgaard J, Deng WQ, Goddard WA (2005) J Phys Chem A 109:493–499
13. Abashkin YG, Erickson JW, Burt SK (2001) J Phys Chem B 105:287–292
14. Rittenhouse RC, Apostoluk WK, Miller JH, Straatsma TP (2003) Proteins 53:667–682
15. Florian J, Goodman MF, Warshel A (2003) J Am Chem Soc 125:8163–8177
16. Köster AM, Calaminici P, Casida ME, Flores-Moreno R, Geudtner G, Goursot A, Heine T, Ipatov A, Janetzko F, Martin del Campo J, Patchkovskii S, Reveles JU, Salahub DR, Vela A (2006) deMon2k
17. Godbout N, Salahub DR, Andzelm J, Wimmer E (1992) Can J Phys 70:560–571
18. Calaminici P, Janetzko F, Köster AM, Mejia-Olvera R, Zuniga-Gutierrez B (2007) J Chem Phys 126:044108
19. Maguire ME, Cowan JA (2002) Biometals 15:203–210
20. Pelletier H, Sawaya MR, Wolfle W, Wilson SH, Kraut J (1996) Biochemistry 35:12742–12761
21. Li Y, Korolev S, Waksman G (1998) EMBO J 17:7514–7525
22. Vassylyev DG, Vassylyeva MN, Zhang JW, Palangat M, Artsimovitch I, Landick R (2007) Nature 448:163–U164