Published in final edited form as:

Adv Protein Chem Struct Biol. 2014 ; 97: 83–113. doi:10.1016/bs.apcsb.2014.10.001.

Applications of Quantum Mechanical/Molecular Mechanical methods to the chemical insertion step of DNA and RNA polymerization

Lalith Perera¹, William A. Beard¹, Lee G. Pedersen², and Samuel H. Wilson^{1,*}

¹Laboratory of Structural Biology, National Institution of Environmental Health Sciences, Research Triangle Park, NC 27709

²Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Abstract

We review theoretical attempts to model the chemical insertion reactions of nucleoside triphosphates catalyzed by the nucleic acid polymerases using combined quantum mechanical/molecular mechanical methodology. Due to an existing excellent database of high resolution x-ray crystal structures, the DNA polymerase β system serves as a useful template for discussion and comparison. The convergence of structures of high quality complexes and continued developments of theoretical techniques suggest a bright future for understanding the global features of nucleic acid polymerization.

I. Introduction

The evolution of modern x-ray crystallography has led to a rapidly increasing wealth of information about the three dimensional structures of both DNA and RNA polymerases (Wu et al. 2014). The resulting high resolution structures, capture intermediates that span the reactants to products path, have created a fertile ground for computational theoreticians to develop, test and apply methods that can expose the finer details of the bonds that form and break. In this article we focus on several nucleic-acid polymerization systems for which sufficient structures exist to reasonably explore the energy requirements of possible pathways.

An overview of the current state of the rapidly changing knowledge of DNA repair enzymes can be found in the more recent review by Sobol (2014) and the more structural review by Wu et al. - (2014). The nucleic acid (DNA and RNA) polymerase chemistry and structures have been systematically discussed in the recent volume edited by Murakami and Trakselis (2014). The basic premise is that good structures of pertinent complexes may serve as starting points for theoretical studies along a reaction path between reactant (pre-chemistry) and product.

*Corresponding Author: wilson5@niehs.nih.gov.

The 2013 Nobel prize for Chemistry recognized the pioneering theoretical developments by M. Karplus, M. Levitt and A. Warshel, the central contribution being the quantum mechanical/molecular mechanical (QM/MM) approach to break a large biological molecular system up into a quantum mechanical core where bonds can break and form and a molecular mechanical system where covalent bonds are not formed or broken. In the latter case, classical force fields can be used to describe the motion and long range electrostatics (Warshel and Levitt, 1976 and Bash et al., 1991). Recent reviews that focus on the development of the method are available (Lin and Truhlar (2007), Senn and Thiel (2009), Lodola and De Vivo (2012), Groenhof (2013) and Konig et al., (2014)).

II. Methods for describing reactive pathways

There are many variations of the QM/MM methodology. Most have to do with the manner in which the boundary between the quantum mechanical and classical mechanical regions is treated. (There is the story of the party of mathematicians that you happen to be attending. One of the group is expounding on a particular idea. The best way to become included in the group, the story goes, is to ask, at a particular point of emphasis in the exposition: “But, what happens at the boundary?”). How to treat the boundary and what happens at the boundary are the essential questions that we face. Let us consider an example of a popular version of QM/MM—the ONIOM method of the Morokuma group (Vreven et al., 2006)). The ONIOM method considers 4 energies— $E(\text{QM, all})$, $E(\text{MM, all})$, $E(\text{QM, Q})$ and $E(\text{MM, Q})$: the QM and MM energies of the complete system and the QM and MM energies of the specified quantum region, Q, respectively. Then the boundary is chosen (in principle) in such a way that the differences between $E(\text{QM, all})$ and $E(\text{MM, all})$ and between $E(\text{QM, Q})$ and $E(\text{MM, Q})$ are equal or nearly so. Then, we would have

$$E(\text{QM, all}) = E(\text{MM, all}) + [E(\text{QM, Q}) - E(\text{MM, Q})] \quad (1)$$

that can be iterated to consistency.

Clearly, if we had sufficient computer resources, we would compute $E(\text{QM, all})$ directly by solving the time dependent Schrodinger equation directly. Lacking that overall ability at the present time for large enzymatic systems, we can, however, if the quantum region size (Q) is modest, determine estimates for the three quantities on the right hand side of equation (1). In the case of the ONIOM method, the boundary is chosen in practice by identifying single bonds in the side chains of residues that project into the proposed catalytic region but that are not expected to take part directly in the bond forming/breaking of the reaction. These bonds are then terminated with hydrogen atoms which are included in the calculation of $E(\text{QM, Q})$ and $E(\text{MM, Q})$. Since $E(\text{QM, Q})$ and $E(\text{MM, Q})$ include the electrostatic fields from atoms outside the Q region (defined as electronic embedding - in practice it becomes necessary to reduce the charges on atoms within several bonds of the Q region). The ONIOM method is flexible in that the critical catalytic region can be treated at a high quantum level, while the immediate surrounding region can also be treated at a lower level quantum region thereby allowing for polarization. A recent application of the ONIOM

(QM/MM) method by Ding et al., (2013) to a photochemically-induced decarboxylation reaction of a green fluorescent protein illustrates the power and flexibility of the method. Developments leading up to this application are considered by Chung et al. (2012).

The major computer modeling packages- (AMBER (Wang, et al., 2004), CHARMM(Brooks, et al., 2009), GROMOS(Scott, et al., 1999), and NAMD(Phillips, et al., 2005)) - all have QM/MM software with varying degrees of flexibility. Individual laboratories have also contributed novel approaches, for example, the pseudobond approach (Zhang et al., 1999), the QTCP approach (Rod and Ryde, 2005) and a high-dimensional string QM/MM free energy method combined with an enhanced-sampling technique (Rosta, Yang & Hummer, 2014).

A particularly insightful and recent review of the application of QM/MM methods to enzymes has been provided by van der Kamp and Mulholland (2013) and a measure of how good certain variants of the method perform vs a full QM calculation can be found in a review by Hu, Soderhjelm and Ryde (2011).

III. DNA Polymerase β

It is probably true that more crystallographic data has been collected on DNA polymerase (Pol) β than any other polymerase (Beard and Wilson, 2014). This 39 kD, DNA gap-filling enzyme has occupied much of the chemical, structural and theoretical effort of our laboratories in recent years. The definitive structure for defining the active site for NTP insertion was published in 2006 (Batra et al., 2006) and was made possible by the use of a nonreactive NTP to impede the reaction. Previous approaches relied on using a dideoxy-terminated DNA substrate that lacked an atom that participated in catalysis; i.e., O3' (Pelletier et al., 1994). The comparative analysis of structures and kinetics of Asp256 (wild-type) along with D256E and D256A variants (Batra, et al., 2013) has established that Asp256 is the catalytic base for triggering the insertion reaction – Specifically, the transfer of a proton from the sugar O3' to the catalytic base initiates the reaction. As described below, this system universally utilized nucleic acid polymerases.

Abashkin et al., (2001): A cluster quantum mechanical calculation at steps along a proposed reaction path

A brave and early attempt at an atomic level understanding the chemistry of the insertion step of Pol β was by Abashkin, Erickson and Burt (2001). This study, while not QM/MM, motivated our later work in that it focused closely on what was known from structural experiments at the time to define a realistic initial state for theory. In this case, the reference crystallographic structure (2.9 Å resolution) was determined by Sawaya et al., (1997). The theoretical study that followed was a set of DFT(double zeta)/QM calculations on an electrostatically neutral cluster of 67 atoms that defined various steps along two postulated reaction paths - (Fig. 1).

A path with an intermediate PO3 was found to have an unreasonably high reaction energy barrier and so an alternate scheme involving a penta-coordinated transition state was explored. In this alternate scheme, the reaction initiates by the proton on the O3' (modeled,

since the x-ray structure used the absence of this group to avert reaction) jumping to the free oxygen on the α phosphate from which it is further transferred, after some adjustments, to an oxygen atom of the β phosphate of the departing pyrophosphate product. The original modeled position on the O3' proton may be an essential element in determining much of what happens downstream. The QM energies of seven structures, including the initial and final minimized complexes were determined. The product state was 19 kcal/mol more stable than the initial state. This is a reasonable result, but mostly based on the 1997 crystal structure. Several modifications to the crystallographic structure were made to arrive at a stable initial QM form which was electrostatically neutral: the necessary O3'H was added, a hydroxide ion was added to the catalytic magnesium ion, a proton was added to Asp190, which helps bridge the two active site magnesium ions, and this proton was placed so as to hydrogen bond with the P α -oxygen that also bridges the magnesium ions. Finally, a hydrogen bond was defined between the side chain of Arg149 and a γ -phosphate oxygen. The catalytic magnesium may be therefore somewhat under-coordinated and the hydrogen added to Asp190 has no structural basis (in fact, it was added for ensuring calculated stability of the initial structure).

Rittenhouse et al. (2003): A cluster quantum mechanical calculation on the pre-chemistry complex of Pol β

Although this study (Rittenhouse et al. 2003) did not map the reaction, it does highlight the details of the pre-chemistry active site. By building possible models of the active site based on the 1997 x-ray crystal structure of Pol β (Sawaya et al. 1997), an active site is proposed in which P α -O, Asp-190 and Asp-192 bridge the two magnesium ions in a largely symmetrical fashion. In addition a water molecule was proposed to be tightly bound to the catalytic magnesium ion. Asp-256 is also proposed to form a tight bond with the catalytic magnesium ion. The three active sites aspartates are chosen to be -1 formal charge while the NTP was set up with a -4 charge. The primer DNA and substrate (NTP) bases were included along with water molecules bound to each magnesium ion. However, the charge balancing arginines and the proton on the metal bridging oxygen present in the Abashkin et al. (2006) study, were not present. Thus the overall charge of the cluster was -3 . In their model, derived from quantum ab initio and DFT optimizations, the O3' proton has an orientation that is poised for transfer to the unbound negatively-charged α -phosphate oxygen.

Radhakrishnan and Schlick (2006): QM/MM study of Pol β (based on pdb=1BPY Sawaya et al. (1997))—This study as well as work by Lin et al. (2006) that follows, appear to be the first works to apply the QM/MM methodology to the Pol β system. The initial structures for the reaction are taken from earlier studies by the authors (Radhakrishnan and Schlick, 2005, 2004) that had focused on subdomain motions associated with NTP binding. Both the correct NTP (Pol β /DNA/dCTP for G:dCTP) and incorrect insertion (Pol β /DNA/dATP for G:dATP) reactions were considered. While based on the 1BPY structure (Sawaya et al. (1997)), modifications were made to the active site so that three aspartates (190, 192, 256) are somewhat involved (precise geometries are not given, particularly for the units that bridge the two magnesium ions). The initial incorrect insertion structure was modeled from the initial correct insertion structure. The QM region including the link hydrogen atoms at the boundaries consists of 86 atoms: 7 atoms from each aspartate

side chain (all assumed to bear -1 charge), 9 atoms from the $O3'H$ region of the primer sugar, 17 atoms from the NTP (assumed to bear -4 charge and terminate with a (link H)- $CH_2-O-P\alpha$ unit), 7 atoms from the Ser 180 side chain, 19 atoms from the Arg 183 side chain ($+1$ charge), two magnesium ions ($+2$ charge) and 7 H_2O molecules (each Mg ion apparently has a water molecule bound in the initial structures). The charges add up to -2 although the charge was indicated to be -1 . When using the 1BPY structure as the starting structural model, $O3'H$ must be added; apparently the location modeled for the proton on $O3'H$ was such that the proton was not hydrogen-bonded to Asp 256. A later high resolution structure suggests this feature (pdb=2FMS; Batra et al. (2006)), where $O3'H$ binds the catalytic magnesium (Mg_{cat}). The QM/MM calculations employed an existing interface between GAMESS-UK (Schmidt et al. 1993) for the QM calculations (6-311G basis) and CHARMM (Brooks et al. 1983) for the MM calculations. The reactions were studied by generating new structures by constrained MD on modeled structures with modified $O3'-P\alpha$ and $O3'-Mg_{cat}$ distances. These new structures were subjected to energy minimization and of these (50 total), 4 intermediates and a final product were found. These forms were additionally subjected to QM/MM dynamics to reach a total of six structure (reactant, four intermediates, and product) for both the correct and incorrect insertion reactions. For both reactions, the first intermediate is found to involve the $-O3'H$ proton jump to a water molecule and, although the paths of the proton after that are somewhat different for the two reactions, it ends up in the product structure on the γ -phosphate of the NTP. The transition barriers for a presumed bipyramidal transition state are estimated to be greater than 18 kcal/mol for the correct insertion and greater than 21 kcal/mol for the incorrect insertion; values that are consistent with those estimated from experimental studies (Beard, et al., 2002; Ahn, et al., 1998). Due to the manner for generating intermediates, a reaction path construction was not possible. Shortly after this study, structural studies were published implying that the correct insertion initial system (pdb=2FMS, Batra et al. (2006)) has the $-O3'-H$ proton strongly hydrogen-bonded to Asp 256 and the incorrect insertion initial system (Batra et al., (2007)) for G:dATP has $O3'$ displaced from Mg_{cat} with a water molecule completing the catalytic metal hydration sphere. Both of these structural findings were accommodated in the next three studies discussed.

Lin et al. (2006): QM/MM study of Pol β (based on pdb=2FMS Batra et al., (2006))—New structures of Pol β indicated that the catalytic metal site can be occupied by sodium or magnesium ions (Batra et al., 2006). Refinement of the latter magnesium structure was at 2.0 Å. The major change from the 1997 lower resolution structure was that the catalytic metal was now octahedrally coordinated, with clear density for all water ligands, Asp 256 and the $O3'$ oxygen of the primer terminus sugar.

After performing cluster calculations (DFT/B3LYP) on several models suggested by the 2FMS structure, and motivated by the earlier DFT/QM study of Abashkin et al., (2001) and by the QM study of Rittenhouse et al., (2003), Lin et al., (2006) were able to map a stable QM/MM reaction path (not shown) that did not involve a proton on the bridging $P\alpha$ oxygen. Arginine residues 183 and 149 also were not included. Asp 256, which is bound to the catalytic metal, serves as the catalytic base for transfer of the $O3'$ proton. Separate QM/MM calculations to determine the most stable position of the $O3'$ proton for the cluster found

that it was located between O3' and a Asp 256 oxygen atom as the central part of a hydrogen bond. The initial geometry of the O3'H proton is consistent with the O3'-OD2-Asp 256 distance (2.81 Å) of the 2FMS structure. Both metals remain six coordinate throughout the reaction. To gain the effect of the entire protein, QM/MM calculations were performed using the ONIOM method (with electrostatic embedding) discussed in Methods (Lin, et al. 2006). The QM region consisted of three water molecules, the NTP with a proton on the γ -phosphate and a link atom after the O3' sugar oxygen, three aspartates residues with link atoms between the β - and α -carbons, and the primer sugar with a link atom at the exit from the sugar ring (Fig. 3) for a total of 64 quantum atoms. In the crystallographic structure (Batra et al., 2006), there is an apparent hydrogen bond linking the O3'-H to an oxygen of Asp-256, and this oxygen, along with O3', is also coordinated to the catalytic metal. This hydrogen bond telegraphs the transfer of the O3' proton to Asp 256 which activates the nucleophile (O3' anion) for attack on the α -phosphate. The cluster study of Abashkin, et al. (2001), based on a less detailed x-ray crystal structure, especially near the catalytic metal, had O3' proton initially transferred to the α -phosphate.

The initial equilibrium structure (before insertion) was obtained by adding protons to the 2FMS structure (Batra et al., 2006). The non-hydrolyzable analog, -2'-deoxy-uridine-5' (α,β)-imido triphosphate, -located in the active site of the x-ray crystal structure was changed to dTTP to facilitate the reaction. Protonation states of amino acids were set at pH=7.0 via <http://propka.chem.uiowa.edu>. All crystal waters were preserved. Water and counter ions were added to provide a box that included 21,367 water molecules, 25 sodium ions and was electrically neutral. The SANDER module of the Amber 8 package (Case et al., 2005) with the Amber ff99 force field (Wang et al., 2004)) was employed for dynamics simulations and minimizations and the particle-mesh Ewald code (Darden et al., 1993; Essmann et al., 1995) was used for long range electrostatic interactions. The ONIOM module implemented in Gaussian 03 (Vreven et al., 2005) was the base QM/MM method with electronic embedding adopted for the study.

The strategy was to map the energy as a function of two variables: the forming O3'-P α bond and the breaking P α -O(P β) bond. Thus, using the scan keyword in Gaussian 03, a map of the energy versus these two variables was generated using the quantum method/basis set of B3LYP/6-31G*. In the early stages of the reaction, the proton on O3' transfers to the Asp 256 carboxylate group with a low barrier of about 3.5 kcal/mole. Once this step occurs, the O3'-P α distances closes until a transition state (with no stable intermediate) occurs. The geometry at the transition state is defined by R-O3'-P α =2.2 Å and R-P α -O(P β)=1.9Å. The transition state barrier above the initial reactant state is 21.5 and 18.0 kcal/mol above the deprotonated intermediate state. These barriers are consistent with experimental estimates of a free energy of activation of 16 kcal/mol (Vande Berg et al., 2001). At the last point along the reaction path, the product state is -5.2 kcal/mol below the initial reactant state and the reactant bond has now closed to R-O3'-P α =1.65 Å and the broken bond has expanded to R-P α -O(P β) to 3.27 Å.

An electrostatic energy decomposition study was undertaken at the transition state to ascertain which amino acids were stabilizing and which were destabilizing. The energy of interaction was taken to be

$$E_{QM/MM}^{esp} = \sum_{\substack{\alpha \in QM \\ \beta \in residue}} \frac{Q_{\alpha} Q_{\beta}}{R_{\alpha\beta}}$$

Then, the difference between the residue's position at the transition state and at the initial deprotonated state (i.e., O3' deprotonated early in the path) is

$$\Delta E = E^{esp}(\text{residue, TS}) - E^{esp}(\text{residue, initial state})$$

The two residues that contributed the most to stabilizing the transition state were Arg-149 and Arg-183 (−4.6 kcal/mol and −7.1 kcal/mol, respectively). These stabilization energies are in accord with their vicinity to the pyrophosphate leaving group in the initial state (modeled from the x-ray crystal structure) and so, being charged themselves, they can modify the response of the breaking bond as the reaction proceeds.

A central observation of this QM/MM study was the remarkable stability of the geometry of the two magnesium ions, both of which interact relatively symmetrically with a O-P α atom, which together form a scaffold upon which the reaction appears to evolve.

Lin et al., (2008): Pol β incorrect insertion QM/MM study (based on pdb=2C2K: Batra et al., 2008)—The appearance of a Pol β mismatch structure (G:dATP) (Batra et al., 2008) provided a new opportunity for understanding of misincorporation at the atomic level. Experimentally, the misincorporation step is experimentally estimated to be > 600 fold slower for G:dATP than G:dCTP (Ahn et al., 1998; Bakhtina et al., 2005), but it does occur and is measurable. Examination of the x-ray crystal structure pdb=2C2K of this mismatch (Batra et al., 2007) suggests why. The O3'H of the primer terminus is no longer coordinating the catalytic metal; it has been displaced by a water molecule. A direct path was tested to determine if O3'H would bond with P α after the O3' proton transferred to the P α free oxygen. The energy barrier was found to be very high for this path (i.e., 48 kcal/mol) suggesting that another path must instead be viable. Another alternative was to instead propose a two step mechanism, in which the intrusive water molecule was synchronously moved from the magnesium coordination at the same time the O3'H group gained coordination. Constrained molecular dynamics was used to force this transfer and provide a prechemistry state that had O3' strongly interacting with the catalytic metal. The resulting structure, found by equilibration molecular dynamics, was very similar to the Lin et al (2006) prechemistry structure. The barrier for this process was found to be about 14 kcal/mol from a B3LYP/6-311G** calculation on a cluster that included key coordinations for the two metals. The O3'-Mg_{cat} distance served as the driving coordinate for the transformation. The idea was that the ground state structure, derived from careful equilibration of the x-ray crystal structure by molecular dynamics, and which had a high reaction barrier, underwent a conformational change (O3'H moves to displace the Mg_{cat} coordinating water) that costs 14 kcal/mol, but which created a prechemistry structure from which the reaction path energetics could be determined from QM/MM procedures as in Lin

et al., (2006). The energy barrier, using QM/MM procedures with the ONIOM method similar to Lin et al., (2006) for the misinsertion step is about the same as found for the correct insertion. The ratio of the insertion rates (correct/incorrect) is 12.5/.019, which can be converted to an energy difference of 3.8 kcal/mol, and then interpreted as the difference between the ground states of G:dCTP vs G:dATP. In this view, for correct insertion, the ground state is the prechemistry state, while for incorrect insertion, the ground state is separated thermodynamically from the prechemistry state by 3.8 kcal/mol and by a non-rate limiting barrier of 14 kcal/mol (water $\leftrightarrow$ O3' switch). The switch must occur before reaction can occur. Electrostatic effects of residues lying outside the quantum region on the transition state energies were found to be similar to those found for the correct insertion study.

Batra et al., (2013): A QM/MM study based on the x-ray crystal structure of the Pol β variant D256E

The two previous studies (Lin et al., 2006, 2008) concluded that the catalytic base for the deprotonation of the -O3'H proton of the primer terminus was the oxygen atom of Asp 256 that is also bonded to the catalytic magnesium ion. In order to test this idea further, two variant structures (Fig. 4) were determined: D256E and D256A (PDB codes: 4JWM and 4JWN, respectively). The D256E structure (pdb=4KWM) is charge-conservative but significantly different than the D256 structure (pdb=2FMS) in two important ways: the carboxylate side chain is not bonded to the catalytic metal as in wild type (the coordination position on the catalytic metal is now occupied by a water molecule) and, Arg 254, which helps anchor the side chain of Arg 256 in place in the wild type, no longer interacts with the 256 position carboxylate. The -O3'H oxygen is, however, still in essentially the same place as for the wild type (pdb=2FMS) structure, approximately 3.5 Å from the Pa of the NTP. Also, for both the D256E and wild type structure, there exists an apparent hydrogen bond between -O3'H and a carboxylate (256 position) oxygen. The change is more drastic for the D256A system—the structure (pdb= 4JWN) does not have the bimetallic active site, but instead has only one magnesium ion which is located in the nucleotide binding position.

The catalytic metal site is empty (Fig. 4) and the -O3'H oxygen is displaced to 4.9 Å from the Pa of the NTP (as compared to about 3.5 Å in the 2FMS wild type structure. Kinetic measurements of k_{pol} of the insertion rate indicate that the rate of insertion for the D256E system was reduced by three orders of magnitude as compared to the wild type while the D256A system had no measurable activity at pH 7.4. Increasing the reaction pH recovered significant activity (~10-fold/pH unit) indicating that the required deprotonation event had a pK above 10. QM/MM reaction paths, using a QM region were similar to those of the Lin et al. (2006) study. In addition, the basis set was now 6-31G*, an extra -CH₂ group was included in the QM region on the aspartates and only the O3'-Pa distance was varied were mapped using the electronic embedding methodology of the GAUSSIAN 09/ONIOM interface (Case et al., 2010; Vereven et al. 2006) for both the wild type (a QM/MM study was included for internal reference) and the D256E variant systems. Because the alignment of O3'-Pa-O β is similar to that of the wild-type system preprocessing to obtain a pre-chemistry state was not necessary. To test the location along the reaction path of the -O3'H deprotonation, two paths were investigated: one where O3' deprotonates early and one

where it de-protonates late; i.e., near the transition state. The former case occurs for the wild type (as in the Linet al. (2006) study) and the latter case is found for the variant form. The transition state barriers are 14 and 21 kcal/mol for the wild type and variant cases, respectively, while the O3'-Pa distance at the transition state is the same for both cases. The higher energy barrier for the variant is consistent with the greatly lowered experimental insertion rate observed for the variant. A charge analysis using the Merz-Kollmann (Besler et al., 1990) option of GAUSSIAN 09 showed that a more effective nucleophile (O3' anion) is developed at the transition state for the wild type as compared to the variant system. An electrostatic energy decomposition analysis comparison between the wild type and D256E variant did not reveal a clear reason as to why the variant transition barrier is higher; however, when the electronic embedding of the ONION procedure was turned off, thereby minimizing the transition state lowering effects of residues outside the quantum region, the barrier increase was much greater for the variant (increased from 21 to 58 kcal/mol) than for the wild-type enzyme (increased from 14 to 42 kcal/mol). Thus the location of the Glu256 side chain in the variant D256E is more destabilizing overall relative the location of the Asp256 side chain in the wild-type system, even though both systems are aligned well for the reaction and both have an apparent hydrogen bond for the -O3'H hydrogen with the carboxylate of the catalytic base. This particular paper is especially interesting because it combines structure determination, kinetic studies and theoretical estimation of a reaction path to investigate the nature of the chemistry step of the Pol β system.

The broader picture: Pol β

Some DNA polymerases such as Pol β undergo subdomain conformational changes when the NTP substrate binds whereas others (e.g., Pol λ) do not (Wu et al., 2014). When the conformational change is present, the question arises as to whether these pre-chemistry events are kinetically and/or thermodynamically influence the overall mechanism. In addition, there are related questions as to what property (or properties) of the system control substrate discrimination (i.e., fidelity). Theoretical viewpoints differ somewhat, as seen by the attempt by Mulholland et al. (2012) to adjudicate a lively discussion between Schlick et al. (2012) and Prasad et al. (2012) about the relative contributions to the mechanism of pre-chemistry conformational changes to the observed barrier for catalysis. Further insight into the mechanism of DNA polymerases, if not resolution of conflicts, was provided by a commentary by Tsai (2014) and the experimental work (Olson et al., 2013) that showed that for the free energy difference between correct and incorrect insertion (averaged over two families and many substrates) was of the order of 5 kcal/mol. This result would imply that the moderate fidelity of the Pol β may simply be due to this large thermodynamic difference, although it does not tell us about the underlying molecular details. Finally, the relationship of Pol β 's mechanism of action to accumulating structural data, kinetics and computation has been summarized recently (Beard and Wilson, 2014).

IV. Application of QM/MM to systems similar to Pol β

Five of the major DNA polymerase families (A, B, C, X and Y) were recently compared structurally for similarities (Wu et al. 2014) with a focus especially on the geometry of the active site of the ternary complexes (polymerase, double stranded DNA and nucleoside

triphosphate) that lead to the chemistry of the formation of the O3'-P α bond and the breaking of the P α -O-P β bond to form pyro-phosphate. Essential elements that appear to span these five families in the active site are the existence of two negatively charged aspartate side-chains that bridge two divalent metal ions that are separated by 3.5–4.0 Å with nothing directly between these highly charged ions, the full NTP unit, the primer terminal -O3'H. Another essential element appears that for optimal function, including Watson-Crick insertion fidelity, the two divalent metals ions should be magnesium ions. Given this background, let us consider three recent QM/MM studies, one of which consider chemical insertion in another member of the X family (Pol λ) and two of which consider chemical insertion in members of the Y family.

a) Pol λ insertion

For Pol λ (X-family DNA polymerase) insertion (Cisneros et al, 2008) is a QM/MM study based on a published x-ray crystal ternary substrate complex structure (pdb=2PFO; Garcia-Diaz et al., 2007) which had many features similar to the structure of Pol β (pdb=2FMS). The x-ray crystal structure (Fig. 5) consisted of the pre-catalytic complex of double strand DNA, Pol λ and a non-hydrolysable NTP (dUMPNPP) served as the starting template for a QM/MM study of the chemical insertion reaction. To prepare the system, the Mn(II) ion in the catalytic metal site was replaced by a magnesium ion and the dUMPNPP was replaced by dUTP, hydrogen atoms were added, system was solvated in a large box of water and equilibrated with a 2 ns PMEMD simulation (Case et al., 2005). All atoms within 30 Å of the catalytic metal were retained for the starting system. The QM/MM calculations were performed with the pseudo-bond method (Zhang, Lee, and Yang (1999); Zhang (2005)) which employed a modified version of Gaussian 03 (Frisch, 2004) with TINKER (Ponder, 1998) to compute energies along the reactions paths studies. Reaction coordinates for the proposed paths were chosen as described (see below) in Cisneros et al. (2008). A quantum mechanical subsystem was chosen that consisted of the NTP through the C5' sugar atom, side chains of Asp490 (the equivalent of Asp256 in Pol β) and Asp427 and Asp429 (Asp190 and 192 in Pol β), the two magnesium ions, part of the primer sugar terminus (excluding the phosphate and C5') and two metal bound water molecules for a total of 72 atoms (Fig. 5). The boundary atoms defining the pseudo-bond locations are the aspartate C α 's, C5' of the primer dC and C4' of the incoming NTP. After protonating the γ -phosphate, the NTP had a formal charge of -3 and the QM region a net charge of -2. The QM method was B3LYP (Becke, 1993; Lee, Yang, Parr, 1988) with a combined basis of 6-31G* for atoms involved in the paths proposed, the 3-21G for the non-reactive atoms and the LANL2DZ pseudo-potential (Wadt and Hay, 1985) was employed for Mn (paths in which the catalytic metal was either Mn²⁺ or Mg²⁺ were investigated; Mg was always the NTP binding metal). An extra diffuse function was included on P α to accommodate P α hybridization changes. The techniques used to produce the reaction path coordinates (Cisneros, 2008) required a product state structure; this was produced from the reactant state using modeling and QM/MM optimization. This non-experimental structure which anchors the product end of the reaction path thereby interjects a degree of uncertainty into the process. Once a given test reaction path is specified, a reaction coordinate can be defined. Unfortunately the equations defining the reaction coordinates for the reaction paths are not stated explicitly. In this study, the several paths (with magnesium in both sites) that were investigated were initiated by i)

transfer of the O3'H proton to Asp 490, the analog of Asp 256 in the Pol β system, ii) the transfer of the O3'H proton to Asp 429, one of the metal bridging aspartates, iii) transfer of the O3'H proton to one of the water molecules bound to the metal ions and iv) transfer of the O3'H proton to the free oxygen on Pa. All of the paths, except the first, which transfers the O3' proton to the non-bridging asp on the catalytic metal, have high energy transition states. The systematic generation of the coordinates along the reaction path (equilibrated experimental reactant structure) to (equilibrated generated *in silico* product structure) permitted the determination of the energy versus reaction coordinate profile for the systems with either magnesium or manganese ions at the catalytic metal site. These profiles for both metal ions gave approximately the same value for the activation energy (~17 kcal/mol) with the suggestion of a weakly bound intermediate between two transition states. It may be that the intermediate seen is an artifact of the constraint imposed by the method of defining the reaction coordinates. The distances between the metals change very little over the path (3.5 Å for Mg-Mg and 3.7 Å for Mn-Mg). Overall, despite the fact that different QM/MM methods were employed, the conclusions about the path of the reaction and magnitude of the activation energy determined, is similar to that found in the QM/MM study for the wild type Pol β system (Batra et al., 2013).

b) Dpo4 NTP insertion

The Dpo4 (Y-family DNA polymerase) insertion reaction (Wang and Schlick, 2008) was a QM/MM study based on the x-ray crystal structure of the Dpo4/DNA complex of 8-oxoG:dCTP (pdb=2ASD: Rechko et al., (2006)). The x-ray crystal structure has calcium ions in the active site and -O3'H is missing by design to stop insertion during structure determination. Thus this structure is somewhat distorted; for instance, the modeled -O3'-Mg_{cat} distance is 5.3 Å (approx. 3 Å too long), the modeled-in O3'-Pa distance is 4.7 Å (approximately 1 Å too long) and the Me-Me distance is approximately 0.8 Å too long. The distorted x-ray crystal structure was then modeled to be similar to the high resolution Pol β structure (Batra et al., 2006; pdb=2FMS), although it is not clear from the paper's descriptions if the coordination state of the NTP-coordinating Mg ion is complete or the distance between the active site metals. The quantum part of the QM/MM was chosen to include the two bridging aspartates, a Mg_{cat} coordinating glutamate, a small part of the primer terminus, two modeled magnesium ions, the incoming dCTP and four water molecules that coordinate Mg_{cat} and dCTP. The total charge on the quantum system is apparently -3 as the dCTP is fully charged (-4). Once the initial pre-chemistry state was established, several possible paths were investigated. The QM/MM procedure was apparently chosen to be similar to the earlier study on Pol β by Radhakrishnan and Schlick (2006). The boundary between QM and MM is handled by a link hydrogen atom. The assumption was that the O3' must initially be deprotonated before the O3'-Pa bond can form and the Pa-O-P β bond can break. Given this, the paths tested were to transfer the proton from O3' to: i) one of the water molecules, ii) to the free oxygen on the Pa, iii) to an oxygen on Glu108, which is in the same position as Asp256 in the Pol β structure, or iv) to one of the bridging aspartates. The best energy path was concluded to be the transfer of the proton to water molecules with a subsequent transfer to the γ -phosphate. Three intermediates are found along the path to products (Pa-O-P β bond broken, O3'-Pa bond formed) with the highest energy intermediate one in which the proton is transferred through

water molecules to the free oxygen on P α . Although a transition state for this transfer is shown (at 20 kcal above the initial state and 5 kcal above the intermediate), a description of how this curve was determined, and subsequent transition states is not given. The overall conclusion is that the rate-limiting step (O3' proton transferred to water to P α oxygen) occurs about 20 kcal above the reactants.

c) Pol κ insertion

Pol κ is also a Y-family polymerase that bypasses certain lesions such as benzopyrene. This study by Lior-Hoffmann et al. (2012) employs the pseudo-atom method (Zhang et al., 1999). A ternary substrate complex X ray crystal structure (pdb=2OH2; Lone et al., 2007) is employed. This structure, which is missing the -O3'H group and does not have a catalytic ion in place, requires significant modeling to obtain a structure suitable for defining an adequate pre-chemistry system. The modeled system involved establishing octahedral coordination at each magnesium ion. A backbone oxygen of Met108 occupies the sixth position of the NTP Mg_{cat} and Glu 199 occupies the same position as is occupied by Asp256 in Pol β . A reaction coordinate driving procedure which proceeds by "stepping along a proposed reaction coordinate and performing energy minimizations with respect to the remaining coordinates" is employed (Zhang et al., 2000). It is not clear what the actual reaction coordinates used are in the paper, but generally they appear to partially describe the transfer of protons. Also, not all of the geometric parameters are defined. The choice as to which atoms to include in the QM part of the QM/MM scheme is somewhat different than for the Dpo4 system. For the Pol κ system, 81 atoms are included in the QM part: the NTP, two water molecules (one of which resides on Mg_{cat}), the two Mg ions, the primer terminal sugar and base and the side chain of Glu199. The two metal ion bridging aspartates and the backbone atoms of Met 108 (which coordinates the Mg_{nuc}) are relegated to the MM subsystem. The NTP is taken to be fully charged (-4) so that the total system charge is -1. Several reaction paths for the de-protonation of the modeled O3'H were tested: i) employing Glu199 as a catalytic base, or ii) employing the free oxygen on P α as the catalytic base. These potential paths were rejected when the protonated species appeared to be unstable for short interspersed QM/MM-MD simulations. The path (reaction paths tested are never precisely defined) that was found to be desirable was iii) a transfer of the O3'H proton to the γ -phosphate through the two waters included in the QM subsystem leading to a stable intermediate, followed by the transfer of this proton to the β -phosphate as the O3'-P α bond formed and the P α -O-P β bond broke. Similar to the Dpo4 system, the initial proton transfer step was found to be rate limiting. The free energy of activation was found to be approximately 11 kcal/mol.

These last two QM/MM simulations discussed (Dpo4 and Pol κ), although using different ways of handling the boundary between QM/MM, share similarities. Both of the reference x-ray crystal structures are significantly distorted, are missing the O3' unit, and do not have a Mg-Mg ion pair at the core of the active site until modeled. Additionally, neither paper gives geometry details about the modeled active site. For instance, it is impossible to deduce from these papers the distances from O3' to the oxygen atoms of the glutamate that occupies the Asp256 position in Pol β in their pre-chemistry structures. If the Lin et al. (2006, 2008) and Batra et al. (2013) QM/MM papers are correct for the mechanism of Pol β

and other similar systems, the reaction of the chemistry insertion is initiated by the transfer of the O3'H proton to the side chain of Asp256 that is nearest O3' and together Asp256 and O3' are bound to the catalytic magnesium. Both the Dpo4 and Pol κ papers consider the possibility of this mechanism: for Dpo4, the transfer of the O3' proton to Glu108 and for Pol κ , the transfer of the O3' proton to Glu199. In both cases, the cause for rejecting this path is stated to be that the transfer state is found unstable. In both cases, however, not enough detail (which carboxylate oxygen for transfer was tested, what was the distance from the O3' to this oxygen in the modeled pre-chemistry structure, what was the precise geometric system tested for stability) is given to be assured that this path (O3'H to glutamate) was thoroughly vetted. This issue becomes important later.

V. RNA Polymerase

The understanding of the mechanism of the synthesis of mRNA was greatly advanced by the publication of a group of structures of RNA Polymerase II by Wang et al. (2006) derived from *S. cerevisiae*. One of these structures, pdb=2E2H, was chosen by Carvalho, Fernandes and Ramos (2011) as the basis of a QM/QM study on the mechanism of extension of mRNA by Pol II. The structure (resolution 3.95 Å) does not have an O3' on the ribose primer and the α subunit of the GTP substrates is not perfectly seated between the two Mg ions (as in the case of higher resolution structures of DNA polymerases). The Mg ions are 3.43 Å apart which is consistent with known DNA polymerase structures. While Asp 485 is firmly attached to the catalytic Mg ion and two aspartates serve as bridges between the two magnesium ions, the resolution is insufficient to observe a complete coordination shell about the magnesium ions. The ONIOM/Gaussian 03 (Vereven et al., 2006) method is used for optimizing trial structures generated to satisfy trial paths chosen for investigation. A QM/QM procedure in which there is a higher level QM (DFT B3LYP/6-31G(d)) applied to inner core atoms and a lower level QM (PM3MM, Stewart, 1989) applied to an outer shell of atoms. The ONIOM/Gaussian 03 programs allow both QM/MM and QM/QM procedures. Final energies of the optimized structures were then computed at the B3LPY/6-311++G(2p, 2d) level. A neutrally charged model of the active site was extracted (Fig. 6) that consisted of parts of 4 aspartates, GTP, the primer terminal ribose, the two Mg ions, parts of a histidine, a lysine and three arginines for a total of 226 atoms. For the inner core, the two Mg ions, GTP, the ribose, histidine and last three side chain atoms of the four aspartates were chosen. There apparently was no water molecules included in the region of quantum calculation. After preparing the models by adding the missing -O3'H unit to the ribose and missing H atoms, and relaxing in neutralized water for 20 ns, four hypothetical reaction pathways were generated. These were: i) HYP1 (the O3' proton jumps to an α phosphate oxygen on the GTP and ends up protonating the pyrophosphate after the bond forming (O3'-P α) and bond breaking (P α -O β)), ii) HYP2 (a hydroxide ion materializes near the -O3'H, deprotonates O3' so that it can attack P α , and the departing pyrophosphate product is stabilized by a proton transfer from a nearby histidine), iii) HYP3 (an OH- group is added on the catalytic magnesium ion initially, which deprotonates the O3-H, while the product pyrophosphate is stabilized by a proton from the histidine and iv) HYP4 (the non-bridging aspartate bound to Mg_{cat} accepts the proton from -O3'H so that the -O3' anion can attack P α and the histidine protonates the leaving pyrophosphate). The energy cost of generating a

hydroxide from bulk water for the HYP2 path was rationalized to be 7.5 kcal/mol by the use of free energy perturbation theory and concentration in bulk considerations. Only HYP2 (path 2) had a low activation energy (~10 kcal/mol), the other three paths had barriers of > 29 kcal/mol. In this path, a hydroxide is created near O3'H, which facilitates deprotonation of the O3'. Then, the positively charged His1085 loses its proton to one of the free oxygens on P β . The consequence of this is the weakening of the P α -O $\alpha\beta$ bond as the P α -O3' bond forms. It is this last step that defines the limiting reaction step in the overall path. If it is the case, as we speculated earlier, higher resolution structures ultimately lead to an initial active site that has what we think are the essential features (O3' in place, an Asp/Glu group at the Asp 256 Pol β position and a water molecule coordinating Mg_{cat}, the P α -O symmetrically (approximate isosceles triangle) bridging the two magnesium ions and there are two metal-bridging aspartates), this mechanism will require reinvestigation. Indeed, in the same paper that pdb=2E2H originates, one can also find pdb=2E2J. Also, in the latter structure, O3' is present (a non-hydrolyzable NTP is used) and it is tightly bound to the Mg_{cat}. In this structure, the aspartate coordinating to the Mg_{cat} is within H-bonding distance (2.52 Å) of the naturally-present O3'. And a later structure (pdb=3S1Q) has appeared (Liu et al. 2011) from the same group, where the aspartate at Mg_{cat} exists and superposition of P α -O and the two magnesium ions with pdb=2FMS is almost perfect even though O3' is missing in the former structure.

Very recently, the insertion pathway for RNA Pol II has been further investigated (Zhang, 2013) in the Salahub lab at the University of Calgary. Several starting systems were considered: i) model A, based on pdb=2E2H (Wang et al, (2006)), ii) model B, based of pdb=2E2J (Wang et al., (2006)), (Fig. 7) iii) models C1 and C2 where, for both of these models, the NTP(GMPCPP) of pdb= 2E2J was changed to the GTP of pdb=2E2H and optimized and the system equilibrated for either 1ns (Model C1) or 12 ns (Model C2). The sizes of the various structures were reduced by fixing all atom coordinates beyond 20 Å of the NTP P α . The quantum subsystem included parts of sidechains of 4 aspartates (481, 483, 485, 837), 3 arginines (446, 766, 1020), three water molecules, the entire GTP or NTP substrate, the ribose of the primer terminus and the two magnesium ions for a total of 144 atoms and charge of -1 for the quantum region. An in-house QM/MM program was employed that utilized hydrogen link atoms at the boundary. The MM part of the calculation was performed with the CHARMM27 force field (McKerell, et al., 1998, Foloppe & Mackerell, 2000) and the QM part of the calculation was performed with the semi-empirical AM1/d-PhoT method (Cui, Gao & York, 2007). The four models were subjected to the defined QM/MM procedure with the apparent conclusion that model C is most appropriate and that direct transfer of the O3'H proton to the α -phosphate will be the lowest energy path to products. We note that a path with initial transfer of the O3'H proton to Asp 485 (the structural equivalent of Asp 256 for Pol β), the side chain carboxylate oxygen of which is 2.52 Å from O3'H in pdb=2E2J) was not investigated.

VI. Thoughts for future QM/MM simulations on ternary substrate complexes of nucleic acid polymerases

Has the experimental structural evidence that has accumulated to date reached the critical amount needed to be able to develop a more unified, consensus view of how the two metal site at the core of polymerase functions? Twenty years ago, this issue was visited with some controversy (Pelletier, 1994; Steitz, 2004) and left unresolved. The diverse mechanisms presented in the various QM/MM applications to polymerases that are presented in this manuscript reflect the lack of a unified view. However, we now believe that a solid case can be made for the “O3'H → initial proton transfer to an active site acidic residue” as the initial step in polymerase activity. For Pol β, this is Asp 256. The data for this suggestion is collected in Table 1.

This data appear consistent with the proposal that if the crystal structure is performed with an non-hydrolyzable NTP substrate, so that O3' is present, and if the active site contains two magnesium ions, then the O3'H group appears to be within a distance characteristic of a strong hydrogen bond in its interaction with the Lewis base occupying the Asp 256 (Pol-β) position. We note that the data in Table 1 spans several nucleic acid families. These observations support the idea that the essential active site for DNA/RNA polymerase activity may involve two closely spaced magnesium ions, an NTP, two metal bridging aspartates, an O3' that coordinates with the catalytic metal and with a Lewis base bound to the catalytic metal. In the higher resolution structures in Table 1, there is also a water molecule bound to the catalytic magnesium. If any of these features is missing and must be modeled, the modeling must be carefully done to restore all essential geometry for function. *One reasonable way to interpret the data in Table 1* is with the assertion that the insertion chemistry is triggered by the proton transfer of the O3' proton through its hydrogen bond to the catalytic base (an aspartate or glutamate). We propose that this jump may take place by quantum mechanical tunneling. Proton tunneling is now a well-established concept for enzyme reactions (Kuznetsov and Ulstrup, 1999; Truhlar et al., 2004; Truhlar, 2009, Hay and Scutton, 2012, Klinman and Kohen, 2013 and Kippenstein, Pande and Truhlar, 2014). Once this jump occurs,” the cat is out of the proverbial bag”. Instantaneously there is a reorganization of electronic charge involving at a minimum the proton, the catalytic metal, the Lewis base and O3'. And, it is likely that the conserved water at Mg_{cat} is involved, perhaps by its induced ionization to lose a proton to a water network with almost simultaneous transfer of the nearby just-added Asp256 proton. It is important to appreciate that there are at least three separated time scales in action here: the widely separated time scales of electrons and heavy nuclei (C, O, N), and the intermediate time scale of protons. The time scale of charge reorganization should be on the order of femtoseconds since it is electronic in nature. It is possible that the proton on the Lewis base at the Asp256 position (Pol β), once transferred from O3', is short lived (it may be transferred through a water network that includes the conserved water molecule on the catalytic metal ion to the region of the NTP near Oαβ). All of these reasonable events will happen with the heavy nuclei essentially static. Instantaneously, we have the Pα-Oαβ bond weakened and the O3'-Pα interaction strengthened. As a result, inversion at Pα begins. Bond forming/breaking happens, perhaps nearly spontaneously. In this view, there is little activation energy of these

steps, presumably the rate controlling step that defines the turnover number is, in fact, the collective energy needed to snap all of the molecular parts in place (or perhaps to clear the active site once the reaction is over). Once in place - with the NTP symmetrically bridging the two metals at Pa, the O3'-Asp256 position hydrogen bond with both O3' and the residue at the 256 position bound to the catalytic metal—the reaction is initiated by the tunneling event between O3' and Asp256. (In the QM/MM simulations on Pol β (Lin et al 2006 and 2008; Batra et al, 2013), the proton was transferred to Asp256 in the slower process of classical barrier crossing and then retained on the Asp256 position Lewis base while the insertion reaction was forced to occur. Tunneling of the proton was not evaluated and a sufficient water network to bridge Mg_{cat} and O $\alpha\beta$ was not in place.) The structural data on the DNA/RNA polymerases (Table 1) thus lead us to a view of polymerase reactivity that may only be fully revealed through the consideration of proton tunneling and multiple time scale events.

VII. Conclusions

All QM/MM studies to date on polymerase reactions, either begin with models that do not correctly account for the necessary strong hydrogen bond at the O3'-Asp/Glu interaction, do not have enough water molecules present to allow for the proton to end up at O $\alpha\beta$ with little energy cost or have quantum sub-systems which are charged, rather than neutral. The ideal pre-calculation crystal structure would have this key hydrogen bond in place, which requires the use of a non-hydrolysable NTP. In this inferred ideal structure, the two magnesium ions would be about 3.5 Å apart, the NTP would be in place, binding the Pa subunit symmetrically to the two metals and two aspartates would bridge the two magnesium ions. The model extracted from this would require three neutralizing positively charged amino acids around the periphery. These would ideally be arginines, which are less resistant to donating protons to the NTP than lysine. The structure of DNA Pol β has three arginine groups ideally located for model building. Tunneling has generally not been considered in the DNA/RNA insertion chemistry reactions to date. Hopefully future work in this area will strive to be described in such a manner as to be reasonably reproducible.

Acknowledgments

This research is supported by Research Project Numbers Z01-ES050158 and Z01-ES050161 to S.H.W and Z01-ES043010 to L.P. by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences. L.G.P. acknowledges NIH grant HL-06350.

Abbreviations List

QM/MM

A procedure that divides a large molecular system into a quantum mechanical part (the smaller part) and a molecular mechanical part (the larger part) with various techniques used to minimize the disruption that occurs at the boundary. The quantum mechanical calculation allows for bond breaking and bond forming.

Pol β	A moderate fidelity member of the X-family of DNA polymerases for which many high resolution crystal structures exist.
NTP	Deoxyribonucleoside or ribonucleoside triphosphate.
DFT	Density Functional Theory. An alternate theoretical procedure for determining energies and structures of molecules. The theory is still in a developmental stage.
Ab initio	An invented term, by Robert Parr and coworkers, which means “from the beginning” in quantum mechanics. The sense in which the term is used is to imply “more accurate” and “from first principles” as opposed to empirical or semi-empirical.
Fidelity	The accuracy of DNA or RNA polymerization and is the reciprocal of the misinsertion frequency
Insertion	The bonding of the nucleotide fragment (NMP) of an NTP to the O3' growing nucleic acid terminus
PME	Particle Mesh Ewald; A fast Fourier transform method to compute the electrostatic energy of a periodic array of charges. (see Darden, York & Pedersen, 1993; Esseman, et al., 1995).
PMEMD	An efficient, modified version of the Sander module of AMBER to perform PME calculations. (written by Bob Duke in the laboratories of Tom Darden and Lee Pedersen)
PM3MM	A semiempirical quantum mechanical code for calculation of properties of molecules. It contains optional corrections for HCON linkages. It is fitted to thermochemical and spectroscopic data. The programs trace to the vision of M.J.S Dewar.
G0X (X=1, 3, 9,..)	Computer packages of Gaussian, Inc, for the calculation of properties of molecules, including ab initio calculation of the energy. The programs trace to the vision of John Pople.
pdb	Protein Data Bank

References

- Abashkin YG, Erickson JW, Burt SK. Quantum Chemical Investigation of Enzymatic Activity of DNA polymerase β . A mechanistic study. *J Phys Chem B*. 2001; 105:287–292.
- Ahn J, Kraynov VS, Zhong X, Werneburg BG, Tsai MD. DNA polymerase β : effects of gapped DNA substrates on dNTP specificity, fidelity, processivity and conformational changes. *Biochem J*. 1998; 331:79–87. [PubMed: 9512464]

- Bakhtina M, Lee S, Wang Y, Dunlap C, Lamarche B, Tsai MD. Use of viscogens, dNT PaS, and rhodium (III) as probes in stopped-flow experiments to obtain new evidence for the mechanism of catalysis by DNA polymerase β . *Biochem*. 2005; 55:5177–5187.
- Bash PA, Field MJ, Davenport RC, Petsko GA, Ringe D, Karplus M. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*. 1991; 30:5826–5832. [PubMed: 2043624]
- Batra VK, Beard WA, Shock DD, Krahn JM, Pedersen LC, Wilson SH. Magnesium-induced assembly of a complete DNA Polymerase catalytic complex. *Structure*. 2006; 14:757–766. [PubMed: 16615916]
- Batra VK, Beard WA, Shock DA, Pedersen LC, Wilson SH. Structure of DNA polymerase β with active-site mismatches suggest a transient abasic site intermediate during misincorporation. *Mol Cell*. 2008; 30:315–324. [PubMed: 18471977]
- Batra VK, Perera L, Lin P, Shock DD, Beard WA, Pedersen LC, Pedersen LG, Wilson SH. Amino Acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. *J Am Chem Soc*. 2013; 135:8078–8088. [PubMed: 23647366]
- Beard WA, Wilson SH. Structure and Mechanism of DNA Polymerase. *Biochemistry*. 2014; 53:2768–2780. [PubMed: 24717170]
- Becke AD. Density-functional thermochemistry. III The role of exact exchange. *J Chem Phys*. 1993; 98:5648–5652.
- Besler BH, Merz KM, Kollman PA. Atomic charges derived from semiempirical methods. *J Comput Chem*. 1990; 4:431–439.
- Biertumpfel C, Zhao Y, Kondo Y, Ramon-Maiques S, Gregory M, Lee JY, Masutani C, Lehmann AR, Hanaoka F, Yang W. Structure and Mechanism of human DNA polymerase ϵ . *Nature*. 2010; 465:1044–1048. [PubMed: 20577208]
- Brooks BR, Brooks CL, Mackerell AD, Nilsson I, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation package. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
- Carvalho ATP, Fernandes PA, Ramos MJ. The catalytic mechanism of RNA Polymerase II. *J Chem Theory Comput*. 2011; 7:1177–1188. [PubMed: 26606364]
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005; 26:1668–1688. [PubMed: 16200636]
- Case, DA., Darden, TA., Cheatham, TE., Simmerling, CL., Wang, J., Duke, RE., et al. *Amber 11*. U. Cal; San Francisco: 2010.
- Chung LW, Hirao H, Li X, Morokuma K. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2012; 2:327–350.
- Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \ln(N)$ method for Ewald sums in large systems. *J Chem Phys*. 1993; 98:10089–10092.
- Ding L, Chung LW, Morokuma K. Reaction mechanism of photoinduced decarboxylation of the photoactivatable free fluorescent protein: An ONIOM(QM/MM) study. *J Phys Chem B*. 2013; 117:1075–1084. [PubMed: 23272644]
- Essmann Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995; 103:8577–8593.
- Foloppe N, Mackerell AD. All-atom empirical force field for Nucleic Acids. I Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comp Chem*. 2000; 21:86–104.
- Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JA., et al. *Gaussian 03*, Revision D.02. Gaussian, Inc; Wallingford, CT: 2004.
- Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. Role of the catalytic metal during polymerization by DNA polymerase λ . *DNA Repair*. 2007; 6:1333–1340. [PubMed: 17475573]
- Groenhof, G. Introduction to QM/MM simulation in Biomolecular Simulations: Methods and Protocols in Molecular Biology. Monticelli, L., Salonen, E., editors. Vol. 924. Springer Science + Business Media; New York, NY: 2013. p. 43–66.
- Hay, s, Scrutton, NS. Good vibrations in enzyme-catalyzed reactions. *Nature Chemistry*. 2012; 4:161–168.

- Hu H, Yang W. Free energies of chemical reactions in solution and in enzymes with ab initio QM/MM methods. *Annu Rev Phys Chem.* 2008; 59:573–601. [PubMed: 18393679]
- Hu L, Soderhjelm P, Ryde U. On the convergence of QM/MM energies. *J Chem Theory Comput.* 2011; 7:761–777. [PubMed: 26596307]
- Klinman JP, Kohen A. Hydrogen tunneling links protein dynamics to enzyme catalysis. *Ann Rev Biochem.* 2013; 82:471–96. [PubMed: 23746260]
- Klippenstein SJ, Pande VS, Truhlar DG. Chemical kinetics and mechanisms in complex systems: A perspective on recent theoretical advances. *J Am Chem Soc.* 2014; 136:538–546.
- Konig G, Hudson PS, Boresch S, Woodcock HL. Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett re-weighting schemes. *J Chem Theory Comput.* 2014; 10:1406–1419. [PubMed: 24803863]
- Kuznetsov AM, Ulstrup J. Proton and hydrogen atom tunneling in hydrolytic and redox enzyme catalysis. *Can J Chem.* 1999; 77:1085–1096.
- Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation formula into a functional of the electron density. *Phys Rev B.* 1988; 37:785.
- Lin H, Truhlar DG. QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc.* 2007; 117:185–199.
- Lin P, Pedersen LC, Batra VK, Beard WA, Wilson SH, Pedersen LG. Energy Analysis of chemistry for correct insertion by DNA polymerase β . *Proc Nat Acad Sci USA.* 2006; 103:13294–13299. [PubMed: 16938895]
- Lin P, Batra VK, Pedersen LC, Beard WA, Wilson SH, Pedersen LG. Incorrect nucleotide insertion at the active site of a G:A mismatch catalyzed by DNA polymerase β . *Proc Nat Acad Sci USA.* 2008; 105:5670–5674. [PubMed: 18391201]
- Lior-Hoffmann L, Wang L, Wang S, Geacintov NE, Broyde S, Zhang Y. Preferred WMSA catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase κ elucidates error-free bypass of a bulky lesion. *Nucl Acid Res.* 2012; 40:9193–9205.
- Liu X, Bushnell DA, Silva DA, Huang X, Kornberg RD. Initiation complex structure and promoter proofreading. *Science.* 2011; 333:633–637. [PubMed: 21798951]
- Lone S, Townson SA, Uljon AN, Johnson RE, Brahma A, Nair DT, Prakash S, Prakash L, Aggarwal AK. Human DNA polymerase κ encircles DNA: implications for mismatch extension and lesion bypass. *Mo Cell.* 2007; 25:601–14.
- MacKerrell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* 1998; 102:3586–3616. [PubMed: 24889800]
- Mullholland AJ, Roitberg AE, Tunon I. Enzyme and catalysis in the mechanism of DNA polymerase. *Theor Chem Acc.* 2012; 131:1286–1288.
- Nam K, Cui Q, Gao J, York DM. Specific reaction parameterization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O and P atoms. *J Chem Theory Comput.* 2007; 3:486–504. [PubMed: 26637030]
- Olson AC, Patro JN, Urban M, Kuchta RD. The energetic difference between synthesis of correct and incorrect base pairs account for highly accurate DNA replication. *J Am Chem Soc.* 2013; 135:1205–1208. [PubMed: 23316816]
- Phillips JC, Braun R, Wang W, Gumbart J, Tajhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
- Ponder, J. Software Tools for Molecular Design, Version 3.6. Washington U: St. Louis; 1998. TINKER.
- Prasad BR, Kamerin SCL, Florian J, Warshel A. Prechemistry barriers and checkpoints do not contribute to fidelity and catalysis as long as they are not rate limiting. *Theor Chem Acc.* 2012; 131:1288–1303.
- Radhakrishnan R, Schlick T. Correct and incorrect nucleotide incorporation pathways in DNA polymerase β . *Biochem Biophys Res Comm.* 2006; 350:521–529. [PubMed: 17022941]

- Radhakrishnan R, Schlick T. Fidelity Discrimination in DNA polymerase β : Differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair. *J Am Chem Soc.* 2005; 127:13245–13252. [PubMed: 16173754]
- Radhakrishnan R, Schlick T. Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase β 's closing. *Proc Nat Acad Sci USA.* 2004; 101:5970–5975. [PubMed: 15069184]
- Rechakoblit O, Malinina L, Cheng Y, Kuryavii V, Broyde S, Geacintov NE, Patel DJ. Stepwise translocation of Dpo4 polymerase during error free bypass of an oxoG lesion. *PLoS Biol.* 2006; 4:e11. [PubMed: 16379496]
- Rittenhouse RC, Apostoluk WK, Miller JH, Straatsma TP. Characterization of the active site of DNA Polymerase β by molecular dynamics and quantum chemical calculations. *Proteins: Structure, Function and Genetics.* 2003; 53:667–682.
- Rod TH, Ryde U. Quantum mechanical free energy barrier in enzymatic reaction. *Phys Rev Lett.* 2005; 94:138302. [PubMed: 15904045]
- Rosta E, Yang W, Hummer G. Calcium inhibition of Ribonuclease H1 two-metal ion catalysis. *J Am Chem Soc.* 2014; 136:3137–3144. [PubMed: 24499076]
- Sawaya MR, Prasad R, Wilson SH, Kraut J, Pelletier H. Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. 1997
- Schlick T, Arora K, Beard WA. Perspective: pre-chemistry conformational changes in DNA polymerase reaction. *Theo Chem Acc.* 2012; 131:1287–1294.
- Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennel J, et al. The Gromos Biomolecular Simulation Package. *J Phys Chem A.* 1999; 103:3596–3607.
- Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Ange Chem (Inter Ed).* 2009; 48:1198–1229.
- Sobol, RW. DNA repair polymerases. In: Murakami, KS., Trakselsis, MA., editors. *Nucleic Acid Polymerases.* Vol. 30. Springer-Verlag; Berlin: 2014. p. 43-84.
- Stewart JJP. Optimization of parameters for semi-empirical methods. II *J Comp Chem.* 1989; 10:221–264.
- Truhlar DG, Gao J, Garcia-Viloca M, Alhambra C, Corchado J, Sanchez ML, Poulsen TD. Ensemble-Averaged variational transition state theory with optimized multidimensional tunneling for enzyme kinetics and other condensed-phase reactions. *Inter J Quant Chem.* 2004; 100:1136–1152.
- Truhlar DG. Tunneling in enzymatic and nonenzymatic hydrogen transfer reactions. *J Phys Org Chem.* 2010; 23:660–676.
- Tsai MD. How DNA Polymerases catalyze DNA replication, repair and mutation. *Biochemistry.* 2014; 53:2749–2751. [PubMed: 24716436]
- Van der Kamp MW, Mulholland AJ. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) methods in computational enzymology. *Biochemistry.* 2013; 52:2708–2728. [PubMed: 23557014]
- Vreven T, Byun KS, Komaromi I, Dapprich S, Montgomery JA, Morokuma K, Frisch MJ. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J Chem Theory Comput.* 2006; 2:815–826. [PubMed: 26626688]
- Wadt WR, Hay PJ. Ab initio effective core potential for molecular calculation—potentials for main group elements Na to Bi. *J Chem Phys.* 1985; 82:284–98.
- Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comp Chem.* 2004; 25:1157–1174. [PubMed: 15116359]
- Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription role of the trigger loop in substrate specificity and catalysis. *Cell.* 2006; 127:941–954. [PubMed: 17129781]
- Wang Y, Schlick T. Quantum Mechanics/Molecular Mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site recognition. *J Am Chem Soc.* 2008; 130:13240–13250. [PubMed: 18785738]
- Warshell A, Levitt M. Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol.* 1976; 103:227–249. [PubMed: 985660]

- Wu S, Beard WA, Pedersen LG, Wilson SH. Structural Comparison of DNA polymerase architecture suggests a nucleotide gate to the polymerase active site. *Chem Rev.* 2014; 114:2759–2774. [PubMed: 24359247]
- Yang W. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase η . *Biochemistry.* 2014; 53:2793–2803. [PubMed: 24716551]
- Zhang, R. PhD thesis. U. Calgary; 2013. Multiscale simulation of mRNA synthesis by RNA polymerase II.
- Zhang Y, Lee T, Yang W. A pseudo-bond approach to combining quantum mechanical and molecular mechanical methods. *J Chem Phys.* 1999; 110:46–54.
- Zhang Y. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical (QM/MM) methods. *J Chem Phys.* 2005; 122:24224.
- Xia S, Wang M, Blaha G, Konigsberg WH, Wang J. Structural insights into complete metal ion coordination from ternary complexes of a B family RB69 DNA polymerase. *Biochemistry.* 2011; 50:9114–9224. [PubMed: 21923197]

Figure 1.

Proposed pre-chemistry molecular cluster model for pol β (pdb=1BPY Sawaya, et al., 1997). The model has many features revealed by higher resolution structures, however, these structures (Batra, et al., 2006) that exhibit an aspartate (residue 256) rather than OH⁻ on the catalytic magnesium. In addition, the structures reveal a conserved water molecule coordinating this metal and further, do not have a hydrogen ion on the metal-bridging oxygen of P $_{\alpha}$. (Abaskin, et al., 2001).

Fig. 2. High resolution structure (pdb=2FMS; Batra et al., 2006) of the pre-chemistry complex of human Pol β . Both magnesium ions are octahedrally-coordinated.

Fig. 3.

Model of the Pol β pre-chemistry active site based on pdb=2FMS (Batra et al, 2006). A water molecule is shown on each magnesium ion. A third water molecule is not shown for clarity. (Lin et al., 2006).

Fig. 4.

X-ray crystal structure for the active site region of Pol β . a) wild-type enzyme: pdb=2FMS (see Fig. 3). b) D256E variant (pdb=WJWN). A water molecule occupies the vacated Asp 256 oxygen position. The alignment of O3'-P α -O β is, however, similar to the wild-type system. The insertion reaction rate is reduced nearly three orders of magnitude. c) D256A variant (pdb=WJWN). The catalytic metal of a) and b) is lost in D256A variant. The insertion reaction is very slow but can be partially recovered at high pH.

Figure 5. High resolution X-ray crystal structure of the pre-chemistry complex of human Pol λ (pdb=2PFO). The catalytic site is occupied by a manganese ion whereas the nucleotide binding site hosts a magnesium ion. Two water molecules coordinating the metal ions are also shown (Garcia-Diaz. et. al., 2007).

Fig. 6.

The 226 atom catalytic core of pdb=2E2H (Wang et al., 2006) chosen for modeling the RNA Pol II insertion reaction with the QM/QM procedure of G03/ONIOM. (Cavalho, et al., 2011). The higher level quantum region is further reduced (not shown) to the triphosphate of the GTP, the RNA ribose, the side chain terminal 3 atoms of each of the four aspartates, the two magnesium ions and the His 1085 side chain.

Figure 7.

The catalytic core of pdb=2E2I chosen by Zhang (2013) for modeling the insertion of a GTP by RNA Pol II.

Table 1

Proposed O3' proton transfer distance to a hydrogen-bonded base oxygen in x-ray crystal structures that have active site magnesium ions, a primer O3', and a water molecule bound to Mg_{cat}. A non-hydrolyzable NTP substrate is employed in all entries to permit inclusion of O3'.

System	Pdb entry	Resolution Å	O3'H-O(Asp/Glu) Å in pdb	AA at "D256 position"
Pol β	2FMS(c)	2.00	2.81	D256
Pol λ	2PFO(c)	2.00	2.61	D490
Pol μ	4M04(c)	1.90	2.81	D418
Pol η	3MR2(c)	1.83	2.67	E116
Pol β(a)	4JWM(c)	2.00	2.60	E256
RB69(b)	3SPY(c)	1.88	2.65	Water-6
RNA Pol II(c)	2E2J (c)	3.50	2.52	D485 RNA/DNA

a) D256E variant: Glu 256 is not bound to the catalytic metal in the x-ray structure but is displaced by a water molecule. A potential hydrogen bond exists between O3' and Glu256. The proton jump from O3' to E256 takes place just before the transition state, whereas this jump for the wild-type enzyme (O3' to Asp256) take place far from the transition state.

b) 4 non-active site mutations (L415A, L561A, S565G, Y567A) were made to wild-type polymerase for structure determination. RB69 is a B family of DNA polymerase. A structure with O3' and two magnesium ions in place has not been solved with the wt protein.

c) **2FMS**: Batra et. al, (2007); **2PFO**: Garcia-Diaz et al., (2007), Mn²⁺ in the catalytic site, Mg²⁺ in the nucleotide binding pocket; **2M04**: Moon et al., (2014); **2MR2**: Biertumpfel et al., (2010); **4JWM**: Batra et al., (2013), **3SPY**: Xia et al., (2011); **2E2J**: Liu et al., (2006). The O3'-Mg_{cat} distance is 2.04 Å, Mg_{cat}-OD1485=3.7 Å, the non-hydrolysable NTP not well seated (O->CH2). The resolution is too poor to determine if Mg_{cat} coordinates a water. It is possible that given the tight O3'-256O bond (2.52 Å), the long 256O- Mg_{cat} distance (2.93 Å) and the short O3'-Mg_{cat} distance (2.04 Å), that the proton on O3' has transferred to a nearby oxygen (aspartate/glutamate) in the crystalline state.