Supporting Material for

Prechemistry Versus Preorganization in DNA Replication Fidelity

B. Ram Prasad and Arieh Warshel*

Department of Chemistry, University of Southern California, SGM 418, 3620 McClintock Avenue, Los Angeles, California 90089

*Corresponding author Email: warshel@usc.edu

I. Computational methods

The EVB method is a mixed quantum mechanics/molecular mechanics (QM/MM) method has been described extensively.[1,2] This method, describes reactions by mixing resonance states (or more precisely diabatic states) that correspond to classical valence-bond structures, which represent the reactant intermediate (or intermediates) and product states. The potential energies of these diabatic states are represented by classical MM force fields of the form:

$$H_{ii} = \varepsilon_i = \alpha_{gas} + U_{intra}(R,Q) + U_{Ss}(R,Q,r,q) + U_{ss}(r,q) \quad (1)$$

Here, ***R*** and ***Q*** represent the atomic coordinates and charges of the diabatic states, and ***r*** and ***q*** are those of the surrounding protein and solvent. α_{gas} is the gas-phase energy of the *i*th diabatic state (where all the fragments are taken to be at infinity), $U_{intra}(R,Q)$ is the intramolecular potential of the solute system (relative to its minimum), $U_{Ss}(R,Q,r,q)$ represents the interaction between the solute (S) atoms and the surrounding (s) solvent and

protein atoms. $U_{ss}(r,q)$ represents the potential energy of the protein/solvent system ("ss" designates surrounding-surrounding). ε_i is given by equation 1 from the diagonal elements of the EVB Hamiltonian (H_{EVB}). The off-diagonal elements of this Hamiltonian, H_{ij}, are assumed to be constant or they can be represented by a simple function, e.g., an exponential function of the distances between the reacting atoms. These H_{ij} elements are assumed to be the same in the gas phase, in solutions and in the proteins. The adiabatic ground-state energy E_g and the corresponding eigenvector C_g are obtained by solving the secular equation:

$$H_{EVB}C_g = E_g C_g \tag{2}$$

The EVB treatment provides a natural picture of intersecting electronic states that is useful for exploring environmental effects on chemical reactions in condensed phases. The ground-state charge distribution of the reacting species ("solute") polarizes the surroundings ("solvent"), and the charges of each resonance structure of the solute then interacts with the polarized solvent.

The EVB approach evaluates the free-energy function Δg that is needed to calculate the activation energy, $\Delta g^{\ddagger}$, by using a free energy perturbation (FEP) /umbrella sampling(US) method which has been described in details elsewhere.[1]

AS discussed in the main text the key power of the EVB is associated with our calibration on the reference solution reaction (Figure 5). Now although the solution surface is concerted we found out that we can obtain very similar results (see below) with a stepwise surface that allow for significantly more efficient calculations and thus more effective sampling. More specifically, in our study, we constructed the EVB surfaces by exploring both the stepwise associative and the concerted path with a single TS. In the associative mechanism, we considered the initial proton transfer to both bulk water and Asp 256. The corresponding EVB resonance structures are depicted in the Figure 4. Both calibrated EVB surfaces reproduce the barrier of the solution surface and can be used for studies in proteins by replacing the solvent with the protein environment (after demonstrating that the two EVB representations produce similar calculated

barriers in the protein). The stepwise associative mechanism is described by seven different VB states (Figure 4A): the reactant state (I), the intermediate containing both the nucleophile and Asp256 are deprotonated (II), the intermediate containing protonated Asp256 and the deprotonated nucleophile (V), the pentacoordinated phosphorane intermediates (III and VI), and the product states (IV and VII). The concerted mechanism is described by three VB states (Figure 4B); initial reactant state (I), deprotonated intermediate (II), and the final product state (III) barriers in the protein).

The modeling of the different mechanistic options is described in details below

II. 1 Associative path

The reaction steps I → II, II → III, and III → IV, forms the complete associative reaction path for the nucleotide transfer reaction in Pol β with the initial PT to the bulk water. Whereas, the reaction proceeds through the steps I → V, V → VI, and VI → VII when the initial PT occurs to Asp256.

II. 1a Proton transfer to the bulk water (I → II)

In order to have a complete reaction free energy profile, we need to calculate the proton transfer activation barrier, $\Delta g^{\ddagger}_{PT}$. However, since this step is not likely to be the rate limiting in the nucleotide incorporation reaction of Pol β, it is sufficient (in most cases) just to calculate the proton transfer reaction energy, ΔG_{PT}, which in turn requires calculating the pK_a shift of the proton donor. The pK_a shift of the ribose 3'OH group in the protein is calculated by using the FEP method and the following relationship[3]

$$\Delta pK_a = \Delta\Delta G^{wat\rightarrow pro}_{sol,OH\rightarrow O^-} / 2.303RT \tag{4}$$

ΔG_{PT} is then calculated as

$$\Delta G_{PT} = 2.303RT(pK_{a,3'OH} - pH) \tag{5}$$

The pK_a of the ribose 3'OH in water is taken as 13.[4] Here $pK_{a,3'OH}=13+\Delta pK_a$ and $pH=7$ are used in above calculations. At present arguably the most rigorous calculation of the PT step in pol β has been performed by the FDFT study reported in ref.[5], as this study involved ab initio calculations with a complete converging free energy mapping. However, as much as proton transfer energy in proteins is concerned we believe that the calibrated EVB

provided by far the most reliable tool[6] as well as analysis of the requirements for converging pK_a calculations by ab-initio QM/MM approaches.[7]

II. 1b Proton transfer to Asp256 (I → V)

The reference reaction for this step was modeled by considering the proton transfer from methanol to acetate ion in water, where we also included the catalytic Mg^{2+} ion. The Mg^{2+} ion was constrained to the corresponding position in the protein by a force constant of 10.0 kcal/mol, while applying a force constant of 1.0 kcal/mol to the rest of the reacting atom in the model solution reaction. However, there are no constraints were used in the protein calculations.

II. 1c Nucleophilic attack (II → III and V → VI)

The modeling of this process in water was carried out for a system consisting of the incoming nucleotide, with both the catalytic and the binding Mg^{2+} ions and the deprotonated 3' primer residue. The simulation conditions used for the reaction in water were identical as those for the DNA-protein-substrate complex as described below. During the simulation,

the geometrical position of both the Mg^{2+} ions were constrained by applying a positional constraint of 3.0 kcal/mol and no other constraints were imposed on the system. In constructing the EVB surfaces for the reference reaction (the attack of O3' primer on incoming nucleotide).The EVB surface was forced to give an activation barrier of 21 kcal/mol in order to reproduce the corresponding experimental estimate and the general features observed by previous ab initio calculaitons.[8] The resulting EVB parameters are summarized in Table 1S. The same EVB parameters were then used without any change in the simulation of the reaction in Pol β. Since the reacting system associated with both the nucleophilic attack reactions (II → III and V → VI) in water and in the protein are identical, we have calibrated only the step that correspond to the II → III transition and the resulting EVB parameters were used for both the corresponding nucleophilic attack reactions in Pol β where the proton is initially transferred to bulk water and where the proton is transferred to Asp256.

The reactions correspond to the departure of pyrophosphate leaving groups, III → IV and VI → VII, were initiated from the final atomic positions and velocities obtained at the end of the corresponding II → III and V → VI steps.

II. 2 Concerted Mechanism

In the concerted mechanism the initial proton transfer step (I → II) (Figure 4B), which is identical to the corresponding I → II reaction step in the associative mechanism (Figure 4A), leads to the formation of the intermediate II (Figure 4B). From here the reaction proceeds in a concerted way forms the final product through a single transition state.

III. The Parameters

The parameters for the stepwise mechanism were taken as the same parameters used in the evaluation of the free energy surface for the nucleotide transfer reaction of T7 polymerase.[8] However, the bonded and nonbonded parameters for the reactant and product states of concerted path, described in Figure 4B, are same as the parameters of the corresponding states of the stepwise associative mechanism. The calibrated EVB surfaces for the solution reactions following the stepwise associative pathway are depicted in Figure 5A and by the concerted pathway in Figure 5B. The barriers and the reaction free energies associated with each step of the solution reaction were reproduced by adjusting the off diagonal element H_{12}

and the gas phase shift α_0 as described in the study of T7 polymerase[8]. The corresponding EVB parameters obtained are given in Table S1.

IV. 1 Initial structures preparation for EVB calculations

The starting structures for the simulations were generated from the x-ray structures of human pol β (PDB: 2FMS for correct nucleotide pair (R) and 3C2M for incorrect nucleotide pair (W)). The structure 2FMS includes two Mg^{2+} ions in complex with a gapped DNA substrate and a nonhydrolyzable deoxyuridine triphosphate (dUTP) opposite a template adenine. In our study we mutated the imide group between P_α and P_β of the dUTP to a phosphoanhydride oxygen. Unlike the previous x-ray structures of pol β (1BPX and 1BPY), 2FMS include primer terminus 3'-OH along with Mg^{2+}. In the x-ray structure, PDB: 3C2M, of incorrect nucleotide comprise DNA polymerase β-DNA-dGTP, the incoming nucleotide dGTP is paired opposite with dATP. However, in this structure, both the Mg^{2+} ions are substituted with Mn^{2+} hence in the current study we replaced them with Mg^{2+} ions. A significant aspect of this structure is, to accommodate the incorrect nucleotide and protein conformation in closed state, the template strand in the vicinity of the catalytic active site is shifted upstream over 3 Å. This

resulted in the repositioning of the coding base and its complementary template base which leads to the rotation of the primer terminus. This rotation moves O3' of the primer terminus away from the α-phosphate of the incoming nucleotide making it inaccessible for the inline nucleophilic attack.[9]

Since the primer terminal position in the ground state (the region observed in the x-ray of W) is not favorable for the direct nucleotide transfer reaction, a prechemistry state, where the primer terminal O3' is bound to the catalytic Mg^{2+} ion, has been generated from its ground state structure using constrained molecular dynamics simulations. In the current study, we also estimated the relative free energy and the activation barrier associated in forming this prechemistry state from the ground state of W system (r(ES(W)).

IV. 2 Preparation of Initial structures for conformational studies

Kinetic and structural studies[16-21] of several polymerases complexes with DNA suggested that the nucleotidyl transfer reaction is associated with large conformational change between the open and closed states. Following DNA

binding, the polymerases incorporate the incoming nucleotide to form an open substrate complex. This complex undergoes a conformational change and reorganizes the active site groups lead to form a closed ternary complex. In this state the nucleotidyl transfer reaction initiates and the nucleotyde is inserted. This complex then undergoes a reverse conformational change, i.e., closed to open, and releases the dissociated pyrophosphate group. It is believed that the conformational changes associated before the nucleotide transfer reaction are key for monitoring DNA synthesis fidelity i.e., binding of the correct nucleotide triggers the conformational closing pathway where as the incorrect nucleotide inhibits or alters the closing pathway.[22] Hence to assess the actual role of the conformational changes on the fidelity of DNA polymerases, it is instructive to estimate the free energy associated with these motions. Since open to close conformational motion is a complex pathway, it might involve several intermediate states. Therefore, all the relevant structural information about the intermediate states is useful. However, due to the lack of the structural information on these states, we have generated the intermediate and the open conformations using targeted molecular dynamics simulations (TMD). The TMD generation of these structures involved the application of a cartesian forces on all the protein heavy atoms, and forcing the system to move between the closed and open

conformations. The overall open to closed conformational change involves a change in the conformation of the α-helix N (residues 253-326). The structural changes in this region are of interest since substrate binding brings this α-helix to close contact with the nascent base pair of the template and incoming dNTP. All the generated intermediate conformations are half way between the closed and open structures. X-ray structure (PDB: 2FMS) was used as the initial closed structure for R, whereas for W, the ground state x-ray structure (PDB: 3C2M) was used. The spatial occupancy of α-helix N in the open, binary nicked complex (PDB: 1BPZ) of Pol β was used as reference in preparing the open ternary complex structures of corresponding R and W systems of Pol β. The fully solvated simulation center and radius were carefully chosen to enable the simulation sphere to accommodate the entire α-helix N and also allow sufficient space for the atoms to move in the direction determined by the TMD. Furthermore, to reduce the problem associated with TMD procedure we carried out long relaxations on the generated structures.

X-ray studies[22] have shown that in the open structure, Arg258 interacts with Asp192. However in the closed structure, this interaction was hampered by the rotation of the phenyl ring of Phe272 causes flipping of Arg258 and

Asp192 away from each other. This allowed residues Glu295 and Tyr296 to form hydrogen bonds with Arg258 and Asp192 to interact strongly with both the Mg^{2+} ions. Hence while generating the intermediate structures, using TMD, starting from the closed, Asp192 and Arg258 are carefully rotated towards each other in several steps. This rotation also includes the flipping of the phenyl ring of Phe272 away from these two groups. At the end in the open structure, Asp192 and Arg258 strongly interact with each other and the phenyl ring of Phe272 was completely rotated away from Asp192 and Arg258. Thus generated intermediate and open structures were subjected to a 20 ps long equilibration initially by imposing the positional constraints on α-helix N region as well as on Asp192, Arg258 and Phe272 residues. The obtained structures were further subjected to another 20 ps long equilibration but without any constraints. These structures were utilized for the systematic evaluation of the free energy associated in moving from open to the closed conformation through the intermediate structures using our renormalizaton approach. In addition, we also carried out EVB calculations for the evaluation of the activation barriers for the chemical steps following the stepwise associative mechanism described in Figure 2 at all the intermediate and open conformations. These calculations were done for both the R and W systems.

V. Simulation conditions and protocol

The simulated DNA-protein-substrate complex was immersed in 18 Å sphere of water molecules subject to the surface-constraint all-atom solvent (SCAAS)[10] type boundary conditions. The geometric center of the reactive atoms is defined as the center of the simulations sphere. Positions of all atoms beyond 24 Å (from the center of solvent sphere) were fixed at their crystallographic positions, and their nonbonded interactions with the atoms within the simulation sphere were turned off. The structure of the simulated system was first relaxed by 20 ps step MD simulation at 300 K with a step size of 1.0 fs. An harmonic position restraints 200 kcal/mol were applied initially to all the protein and DNA atoms. This allowed the water molecules to adjust in and around the protein environment. The initial simulation was then followed by another 20 ps step simulation applying the 200 Kcal/mol harmonic constraint only on the protein atoms forming the first coordination shell of both the Mg^{2+} ions. The harmonic constraint on these atoms was gradually decreased in next twenty 10 ps step simulations. Finally the complex system was equilibrated for another 20 ps without any constraints with the step size 1.0 fs at 300 K. This served as a starting structure for the FEP calculations.

The EVB and FEP calculations were done with the LRF long rang treatment (which is equivalent to having no cutoff for the electrostatic interactions. All the simulations were performed using the MOLARIS simulation package.[11] The nucleic acid bases were represented by AMBER charges[12] and ENZYMIX vdW parameters (after validation that these parameters give reliable solvation energies and base-pairing energies and structures). A single center model with the vdW parameters r^* = 1.30 and ε = 0.06 kcal/mol[8] was used for the Mg^{2+} ions.

In taking into account the protein ionized residues we used our Monte Carlo (MC) approach[13,14] and evaluated the effect of the ionizable group beyond the first solvation shell by using a macroscopic treatment with a large dielectric constant for charge-charge interaction.[15]

References

1. Warshel A. Computer Modeling of Chemical Reactions in Enzymes and Solutions. New York: John Wiley & Sons; 1997.

2. Hwang J-K, King G, Creighton S, Warshel A. Simulation of Free Energy Relationships and Dynamics of S_N2 Reactions in Aqueous Solution. J Am Chem Soc 1988;110(16):5297-5311.

3. Warshel A. Calculations of Enzymic Reactions: Calculations of pK_a, Proton Transfer Reactions, and General Acid Catalysis Reactions in Enzymes. Biochemistry 1981;20:3167-3177.

4. Izatt RM, Rytting JH, Hansen LD, Christensen JJ. Thermodynamics Of Proton Dissociation In Dilute Aqueous Solution.V. An Entropy Titration Study Of Adenosine Pentoses Hexoses And Related Compounds. J Am Chem Soc 1966;88(12):2641-2645.

5. Xiang Y, Warshel A. Quantifying Free Energy Profiles of Proton Transfer Reactions in Solution and Proteins by Using a Diabatic FDFT Mapping. J Phys Chem B 2008;112:1007–1015.

6. Braun-Sand S, Sharma PK, TChu Z, Pisliakov AV, Warshel A. The energetics of the primary proton transfer in bacteriorhodopsin revisited: it is a sequential light-induced charge separation after all. Biochim Biophys Acta 2008;1777:441-452.

7. Kamerlin SCL, Haranczyk M, Warshel A. Progress in Ab Initio QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pKa, Redox Reactions and Solvation Free Energies. J Phys Chem B 2009;113:1253–1272.

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

9. Batra VK, Beard WA, Shock DD, Pedersen LC, Wilson SH. Structures of DNA Polymerase β with active-site mismatches suggest

a transient abasic site intermediate during misincorporation. Cell 2008;30:315-324.

10. King G, Warshel A. A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. J Chem Phys 1989;91(6):3647-3661.

11. Lee FS, Chu ZT, Warshel A. Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the POLARIS and ENZYMIX Programs. J Comp Chem 1993;14:161-185.

12. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Jr., Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation Of Proteins, Nucleic-Acids, And Organic-Molecules. J Am Chem Soc 1995;117(19):5179-5197.

13. Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain AV, Warshel A. Multiscale simulations of protein landscapes: Using coarse-grained models as reference potentials to full explicit models. Proteins: Struct, Funct, Genet 2010;78:1212-1227.

14. Rychkova A, Vicatos S, Warshel A. On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. Proc Natl Acad Sci U S A 2010;107:17598–17603.

15. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models. Proteins: Struct, Funct, Genet 2001;44:400-417.

16. Ahn J, Werneburg BG, Tsai M-D. DNA Polymerase β: Structure−Fidelity Relationship from Pre-Steady-State Kinetic Analyses of All Possible Correct and Incorrect Base Pairs for Wild Type and R283A Mutant. Biochemistry 1997;36:1100-1111-1107.

17. Ahn J, Kraynov VS, Zhong X, Werneburg BG, Tsai MD. DNA polymerase beta: effects of gapped DNA substrates on dNTP specificity, fidelity, processivity and conformational changes. Biochem J 1998;331:79-87.

18. Berg BJV, Beard WA, Wilson SH. DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase beta. Implication for the identity of the rate-limiting conformational change. J Biol Chem 2001;276:3408-3416.

19. Shah AM, Li SX, Anderson KS, Sweasy JB. Y265H mutator mutant of DNA polymerase beta. Proper teometric alignment is critical for fidelity. J Biol Chem 2001;276:10824-10831.

20. Suo Z, Johnson KA. Selective Inhibition of HIV-1 Reverse Transcriptase by an Antiviral Inhibitor, (R)-9-(2-Phosphonylmethoxypropyl)adenine. J Biol Chem 1998;273:27250-27258.

21. Kraynov VS, Werneburg BG, Zhong X, Lee H, Ahn J, -DTsai M. DNA polymerase beta: analysis of the contributions of tyrosine-271 and asparagine-279 to substrate specificity and fidelity of DNA replication by pre-steady-state kinetics. Biochem J 1997;323:103-111.

22. Sawaya MR, Prasad R, Wilson SH, Kraut J, Pelletier H. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry 1997;36(37):11205-11215.

Table S1. Activation and reaction free energies and EVB parameters for reference reactions in aqueous solution.

System	Reaction Step	Reaction Type[c]	Reaction Mechanism	ΔG_0 (kcal/mol)	$\Delta g^{\neq}$ (kcal/mol)	H_{12}[d] (kcal/mol)	α_0[d] (kcal/mol)
R	I → II[a]	PT	Bulk	8.2[e]			
	II → III[a,f]	Nuka	Bulk	18.5	20.1	180.0	145.0
	III→IV[a,g]	DepLg	Bulk	-24.9	2.6	140.0	68.0
	I → V[a]	PT	Asp	12.5	14.4	60.0	55.0
	V →VI[a,f]	NucA	Asp	18.5	20.1	180.0	145.0
	VI→VII[a,g]	DepLg	Asp	-24.9	2.6	140.0	68.0
	I → II[b]	PT	Bulk	8.2[e]			
	II → III[b]		Bulk	-6.4	20.1	160.0	-14.0
W	I → II[a]	PT	Bulk	8.2[e]			
	II → III[a]	NucA	Bulk	18.5	20.1	164.0	148.0
	III → IV[a]	DepLg	Bulk	-24.9	2.6	130.0	68.0
	I → II[b]	PT	Bulk	8.2[e]			
	II → III[b]		Bulk	-6.4	20.1		

[a]Notation according to the Figure 4A. [b]Notation according to the Figure 4B. [c]Proton transfer (PT), O3' attack on P_α (NucA), departure of the leaving group (DepLg). [d]H_{12} denotes the off-diagonal element of the EVB Hamiltonian. α_0 denotes the gas phase shift of the product state. [e]Experimental reaction free energy determined for pK_a (donor) = 13.0 and solution pH 7.0. [f,g]The energies and EVB parameters of these solution reactions are assumed to be identical.