J|A|C|S
ARTICLES
Published on Web 06/13/2003

Computer Simulation of the Chemical Catalysis of DNA Polymerases: Discriminating between Alternative Nucleotide Insertion Mechanisms for T7 DNA Polymerase

Jan Florián,*,† Myron F. Goodman,‡,§ and Arieh Warshel‡

Department of Chemistry, Loyola University Chicago, Chicago, Illinois 60626, and Department of Chemistry and Department of Biological Sciences, Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, California 90089-1062

Received October 17, 2002; E-mail: florian@usc.edu

Abstract: Understanding the chemical step in the catalytic reaction of DNA polymerases is essential for elucidating the molecular basis of the fidelity of DNA replication. The present work evaluates the free energy surface for the nucleotide transfer reaction of T7 polymerase by free energy perturbation/empirical valence bond (FEP/EVB) calculations. A key aspect of the enzyme simulation is a comparison of enzymatic free energy profiles with the corresponding reference reactions in water using the same computational methodology, thereby enabling a quantitative estimate for the free energy of the nucleotide insertion reaction. The reaction is driven by the FEP/EVB methodology between valence bond structures representing the reactant, pentacovalent intermediate, and the product states. This pathway corresponds to three microscopic chemical steps, deprotonation of the attacking group, a nucleophilic attack on the P_α atom of the dNTP substrate, and departure of the leaving group. Three different mechanisms for the first microscopic step, the generation of the RO^- nucleophile from the 3′-OH hydroxyl of the primer, are examined: (i) proton transfer to the bulk solvent, (ii) proton transfer to one of the ionic oxygens of the P_α phosphate group, and (iii) proton transfer to the ionized Asp654 residue. The most favorable reaction mechanism in T7 pol is predicted to involve the proton transfer to Asp654. This finding sheds light on the long standing issue of the actual role of conserved aspartates. The structural preorganization that helps to catalyze the reaction is also considered and analyzed. The overall calculated mechanism consists of three subsequent steps with a similar activation free energy of about 12 kcal/mol. The similarity of the activation barriers of the three microscopic chemical steps indicates that the T7 polymerase may select against the incorrect dNTP substrate by raising any of these barriers. The relative height of these barriers comparing right and wrong dNTP substrates should therefore be a primary focus of future computational studies of the fidelity of DNA polymerases.

1. Introduction

DNA polymerases are central components of the molecular machinery that controls genetics, heredity, and DNA damage repair. Recent observations of crystal structures of the protein/DNA/dNTP substrate ternary complexes of several DNA polymerases provide an opportunity to determine a mechanism for the DNA polymerization reaction. However, the average geometry of the ground state (GS) of a reacting system as seen from X-ray data lacks information about the transition state (TS) region that is required for elucidating mechanistic details. Because of the short lifetimes of TS structures, this information is difficult to obtain by direct spectroscopic observation. Fortunately, if the GS structure is known, the probability of reaching the TS region can be determined in principle using quantum mechanical/molecular mechanical computer simulations. This probability can then be related to the observed reaction kinetics via transition state theory, thus providing an ultimate structure−function relationship.

The structural origin of DNA replication fidelity has been the subject of extensive research.[1−3] Among DNA polymerases that have been crystalized as ternary complexes, the DNA polymerase from the bacteriophage T7 (T7 pol) exhibits the largest replication fidelity. Previously, we have used computer simulations to evaluate the binding free energy differences to form correct and incorrect base pairs between the templating and dNTP bases in T7 pol.[4] If one of the microscopic chemical steps is rate limiting for the overall enzymatic nucleotide insertion reaction, then the total polymerase fidelity would be obtained by augmenting these binding free energy differences with the calculated differences in activation barriers of the

† Department of Chemistry, Loyola University Chicago.
‡ Department of Chemistry, University of Southern California.
§ Department of Biological Sciences, Hedco Molecular Biology Laboratories, University of Southern California.

(1) Steitz, T. A. *J. Biol. Chem.* **1999**, *274*, 17395−17398.
(2) Doublie, S.; Sawaya, M. R.; Ellenberger, T. *Structure* **1999**, *7*, R31−R35.
(3) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191−219.
(4) Florián, J.; Warshel, A.; Goodman, M. F. *J. Phys. Chem. B* **2002**, *106*, 5754−5760.

chemical step.[5] Therefore, theoretical analysis of the activation barriers for the formation of the new phosphodiester bond to the correct and incorrect nucleotides in the polymerase active site would be instrumental for elucidating sources of the high replication fidelity achieved by T7 pol.

First, however, we need to establish a detailed catalytic mechanism that yields the observed rate enhancement when the rate constant in the enzyme (k_{cat}) is compared to that of the corresponding reaction in aqueous solution (k_w). This alone is a formidable task that requires comparing activation free energies of the rate-limiting steps for several alternative mechanisms that are intuitively selected by the inspection of the observed X-ray structure of the ternary complex. Within the realm of enzymes that catalyze the cleavage/formation of PO bonds, such comparisons have been previously provided for the reactions catalyzed by the exonuclease site of the Klenow fragment of the polymerase I,[6] rasp21 GTPase,[7,8] ribonuclease A,[9] and tyrosine phosphatase.[10,11] These studies were carried out using the empirical valence bond/free energy perturbation (EVB/FEP) calculation,[12] which represents a powerful tool in the arsenal of the computational biochemistry. For example, the EVB/MD analysis of rasp21[7] excluded the glutamine as a base mechanism, which had been commonly believed to be the mechanism of GTP hydrolysis in rasp21. This prediction was later confirmed by protein engineering experiments.[13]

In this paper, we carry out EVB/MD calculations of the activation and reaction free energies along three different pathways for the associative nucleotidyl transfer in T7 pol.

All energies are evaluated for the correct dGTP substrate, which forms a Watson−Crick base pair with a templating dC nucleotide. The investigated pathways for the chemical step in T7 pol share three microscopic steps: the deprotonation of the C3′−OH group of the 3′-terminal deoxyribose of the primer strand (step i); the nucleophilic attack of the C3′−O^- group on the P_α atom (step ii); and the subsequent scission of the $P_\alpha O$ bond of the resulting phosphorane intermediate (step iii). Step iii leads to the formation of inorganic pyrophosphate.

The conformation changes that occur before and after the chemical steps are not examined (for discussion, see section 4). Although it is still unclear if the phosphorane represents a true intermediate, the overall associative (S_N2) character of the reaction has been supported by the geometry of the reactant and product structures,[14] by isotope experiments that showed that the reaction proceeds via the inversion of configuration on the phosphorus atom,[15,16] as well as by ab initio calculations.[17] However, these methods were not able to reveal the nature of the deprotonation of the C3′−OH group that must precede or accompany the nucleophilic attack step. The difficulty reflects in part the fact that the crystal structures of the ternary complex of T7 pol (and of other polymerase complexes) lack the 3′OH hydroxyl group on the 3′-end of the primer.

There are three mechanistic options that should be considered for the initial deprotonation step (step i): proton transfer to Asp654, to one of the nonbridging oxygens of the α-phosphate, and to an OH^- ion penetrating from the bulk solvent. The energetics of these three options, as well as the subsequent reaction steps leading to the product state, are examined in this paper by EVB/FEP calculations. The EVB method[18−20] is a coupled quantum mechanical/molecular mechanical (QM/MM) approach that uses parameters of the quantum Hamiltonian that were adjusted to reproduce known energetics of the model reaction in aqueous solution. Thus, after providing general principles of the EVB method and details of the MD protocol in section 2 (Methods), we present the energetics of the model phosphodiester hydrolysis reaction in aqueous solution and the parametrization of the EVB Hamiltonian in section 3.2. The results of our MD simulations of the ground state of the T7 pol·DNA·dGTP ternary complex are presented in section 3.1. The results for the reaction in the protein active site and the calculated structure of the transition state (TS) of the rate-limiting step are given in sections 3.3 and 3.4. An essential aspect of these calculations is the comparison of enzymatic free energy profiles with the corresponding reference reactions in water using the same computational methodology.

2. Methods

The links between the enzyme structure and its catalytic function presented in this study were derived using the following methodological and conceptual approaches:

(i) Transition state theory.

(ii) Analysis of the observed kinetics and thermodynamics for the proton transfer and the phosphodiester hydrolysis reactions in aqueous solution.

(iii) Ab initio quantum chemical calculations of some aspects of the reactions in solution that are difficult to resolve experimentally.

(iv) Computational analysis of the *differences* in the energetics of identical reactions occurring in water and DNA polymerase, using empirical valence bond (EVB) and free energy perturbation (FEP) calculations.

Further explanation of points (i), (iii), and (iv) is given in the following.

2.1. Transition State Theory. The evaluation of the kinetics of enzymatic reactions, as done in this study, relies on the validity of the classical transition state theory (TST).[21−23] The TST expresses the rate constant for a reaction along a reaction coordinate x as

$$k_{TST} = \frac{k_B T}{h} \exp\{-\Delta g(x^{\ddagger})\beta\} \quad (1)$$

where k_B is the Boltzmann constant, T is the absolute temper-

(5) Florián, J.; Goodman, M. F.; Warshel, A. *Biopolymers*, **2003**, *68*, 286−299.
(6) Fothergill, M.; Goodman, M. F.; Petruska, J.; Warshel, A. *J. A. Chem. Soc.* **1995**, *117*, 11619−11627.
(7) Langen, R.; Schweins, T.; Warshel, A. *Biochemistry* **1992**, *31*, 8691−8696.
(8) Glennon, T. M.; Villà, J.; Warshel, A. *Biochemistry* **2000**, *39*, 9641−9651.
(9) Glennon, T. M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 10234−10247.
(10) Hansson, T.; Nordlund, P.; Aqvist, J. *J. Mol. Biol.* **1997**, *265*, 118−127.
(11) Kolmodin, K.; Nordlund, P.; Aqvist, J. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 370−379.
(12) Warshel, A.; Russell, S. T.; Sussman, F. *Israel J. Chem.* **1986**, *27*, 217.
(13) Chung, H. H.; Benson, D. R.; Schultz, P. G. *Science* **1993**, *259*, 806−809.
(14) Pelletier, H.; Sawaya, M. R.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1891−1903.
(15) Burgers, P. M. J.; Eckstein, F. *J. Biol. Chem.* **1979**, *254*, 6889−6893.
(16) Brody, R. S.; Adler, S.; Modrich, P.; Stec, W. J.; Leznikowski, Z. J.; Frey, P. A. *Biochemistry* **1982**, *21*, 2570−2572.
(17) Abashkin, Y. G.; Erickson, J. W.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 287−292.
(18) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(19) Warshel, A.; Sussman, F.; Hwang, J.-K. *J. Mol. Biol.* **1988**, *201*, 139.
(20) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(21) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875−894.
(22) Eyring, H. *Chem. Rev.* **1935**, *17*, 65−77.
(23) Glasstone, S.; Laidler, K. J.; Eyring, H. *The Theory of Rate Processes*; McGraw-Hill: New York, 1941.

ature, h is Planck's constant, $\beta = (k_B T)^{-1}$, $x^{\ddagger}$ is the value of x at the transition state, and $\Delta g(x)$ is the free energy function in the multidimensional configuration space[20]

$$\Delta g(x) = -\beta^{-1} \ln\{\int \exp[-E(x,s_\perp)\beta]\, ds_\perp / \int \exp[-E(\langle x\rangle_R, s_\perp)\beta]\, ds_\perp\} \quad (2)$$

In eq 2, the variable of integration $s_\perp$ runs over all degrees of freedom orthogonal to the reaction coordinate, $E(x,s_\perp)$ is the energy of configuration $(x,s_\perp)$, and $\langle x\rangle_R$ is the most probable value of x in the reactant state. The factor $\{-\Delta g^{\ddagger}\beta\}$ in eq 1 expresses the probability of finding the reactants in the transition state region of the configuration space.

The insertion of the calculated activation barriers for a reaction step in an enzyme active site ($\Delta g^{\ddagger}_{cat}$) into eq 1 yields the first-order catalytic rate constant k_{cat} (s^{-1}). Similarly, the calculated activation barrier for the reaction occurring in water ($\Delta g^{\ddagger}_{w}$) yields the rate constant for the corresponding reaction in aqueous solution (k_w). Practically, the reactions in water are calculated for reactants that are already in a contact distance, that is, in a common solvent "cage".[24] Thus, the calculated quantity is $\Delta g^{\ddagger}_{cage}$ rather than $\Delta g^{\ddagger}_{w}$. These two activation barriers are identical for unimolecular reactions or for bimolecular reactions when one of the reactants is water. For bimolecular reactions of 1 M reactants in solution, $\Delta g^{\ddagger}_{w}$ (kcal/mol) is related to $\Delta g^{\ddagger}_{cage}$ by the approximate relationship[20]

$$\Delta g^{\ddagger}_{w} \approx \Delta g^{\ddagger}_{cage} + 2.4 \text{ kcal/mol} \quad (3)$$

The rate constant of the TST should in principle be corrected by a transmission factor F, which reflects the true dynamical effects. However, it is now widely accepted (see ref 25 and references therein) that the magnitude of F is close to unity in reactions with activation barriers larger than 10 kcal/mol.

2.2. Ab Initio Calculations of the Energetics of the Solution Reactions. To assess the catalytic effect of the enzyme, it is essential to analyze the free energy of the corresponding solution reaction. Thus we studied the free energy profile for the alkaline hydrolysis of dimethyl phosphate monoanion (DMP^-) in aqueous solution using a combination of gas-phase ab initio calculations at the MP2/6-31+G**//HF/6-31G* level with calculations of the solute polarization and solvation free energies using polarized continuum (PCM)[26,27] and Langevin dipoles (LD)[28,29] solvation models, respectively. Although an initial estimate of geometries of the stationary points were obtained by the geometry optimizations in the gas phase rather than in solution, the actual profile in solution did not entirely rely on the gas-phase geometries but rather on the series of single-point solution calculations in the vicinity of the gas-phase stationary points. The calculations were carried out using Gaussian[30] and Chemsol 1.0 programs. For further details and discussion of the methods, see refs 24 and 31.

2.3. General Concepts of EVB/FEP Calculations of the Activation and Reaction Free Energies. Clearly, to relate the activation barrier of an enzymatic reaction to the structure of the enzyme−substrate complex, it is necessary to use an efficient computational method. This method should be able to capture the energetics of bond breaking and forming events while being fast enough to allow sampling of many configurations of a protein−solvent−substrate system. At present, the empirical valence bond (EVB) approach[18,32−35] appears to be the method that suits best the previously mentioned criteria.

The EVB method describes a reaction in terms of mixtures of diabatic states that correspond to classical valence-bond (VB) structures in the system that is divided between the reacting part (S) and the surrounding protein and solvent (s). First, the potential energy of the ith diabatic state is calculated

$$\epsilon_i = U^i_S(\boldsymbol{R},\boldsymbol{Q}^i) + \sum_{m\in S} V^i_m(\boldsymbol{R},\boldsymbol{Q}^i,\boldsymbol{r},\boldsymbol{q}) + E_s(\boldsymbol{r},\boldsymbol{q}) + \alpha^i_S \quad (4)$$

where $\boldsymbol{R}$ and $\boldsymbol{Q}$ represent the atomic coordinates and charges of S, $\boldsymbol{r}$ and $\boldsymbol{q}$ are the coordinates and charges of s, U^i_S is the intramolecular energy of S in its ith VB state, V^i_m represents nonbonded (van der Waals and electrostatic) interactions of the atom m of S with all the surrounding atoms of s, E_s represents the internal potential energy of the surroundings, and α_i is the gas-phase energy of S when its individual fragments are separated to infinity. Because the charges of S and s regions interact, $E_s(\boldsymbol{r},\boldsymbol{q})$ depends implicitly on $\boldsymbol{R}$ and $\boldsymbol{Q}$, and the $U^i_S(\boldsymbol{R},\boldsymbol{Q}^i)$ energy depends implicitly on $\boldsymbol{r}$ and $\boldsymbol{q}$. The energy functions in eq 4 belong to the category of molecular mechanics force fields, with the exception that some bonds of S are modeled by Morse rather than harmonic potentials. The actual EVB and force field parameters used in this study are presented in section 2.4.

The ϵ_i energies, which are given by eq 4, form the diagonal elements of the EVB Hamiltonian. The off-diagonal elements (H_{ij}) are represented as constants or simple exponential functions of the distances between the reacting atoms. The ground state energy, E_g, obtained by diagonalizing the EVB Hamiltonian provides the adiabatic potential surface for the reaction. The natural picture of intersecting electronic states in the EVB treatment is particularly useful for exploring environmental effects on chemical reactions.[18,33] The ground state charge distribution of S polarizes the surroundings ("solvent"), and the charges of each resonance structure of S then interact with the polarized solvent,[19] yielding new diabatic energy surfaces and, after the diagonalization procedure, a new polarized ground state surface. This coupling enables the EVB model to capture the effect of the solvent on quantum-mechanical mixing of ionized and covalent states of the reacting part. The reliability of the EVB treatment has been demonstrated by comparing the calculated electrostatic free energies to corresponding ab initio quantum-mechanical values.[36]

In principle, very long ($\sim 10^1$ s) molecular dynamics trajectories on the EVB surface of the reactant state could provide the free energy function Δg that is needed to calculate the activation energy $\Delta g^{\ddagger}$ (see eqs 1 and 2). However, because

(24) Strajbl, M.; Florian, J.; Warshel, A. *J. Am. Chem. Soc.* **2000**, *122*, 5354−5366.
(25) Warshel, A.; Parson, W. W. *Q. Rev. Biophys.* **2001**, *34*, 563−679.
(26) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, *55*, 117−129.
(27) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
(28) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.
(29) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 10282−10288.
(30) *Gaussian 94*, revision D.2; Gaussian, Inc.: Pittsburgh, PA, 1995.
(31) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(32) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297−5311.
(33) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.
(34) Florián, J. *J. Phys. Chem. A* **2002**, *106*, 5046−5047.
(35) Jensen, F.; Norrby, P. O. *Theor. Chem. Acc.* **2003**, *109*, 1−7.
(36) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.

required simulation times are at present completely beyond our reach, it is necessary to calculate $\Delta g^{\ddagger}$ using much shorter trajectories ($\sim 10^{-9}$ s) during which the system is driven from the reactant to the product state with the help of a modified potential surface.[32] This can be done using the FEP formulation.[32,37−40] In the simple case of a reaction that is dominated by two (reactant and product) valence-bond states, this so-called "mapping" potential (ϵ_m) can be written as a linear combination of the diabatic reactant and product potentials (ϵ_R and ϵ_P):

$$\epsilon_n = (1 - \eta_n)\epsilon_R + \eta_n\epsilon_P \quad (5)$$

with $0 \le \eta_n \le 1$. If η is changed from 0 to 1 in fixed increments, potentials with one of the intermediate values of η will force the system to fluctuate near the TS. This mapping potential may not provide an accurate representation of the ground state surface (E_g) in the region of the TS, because eq 5 does not diagonalize the Hamiltonian of the system, but it does not need to do so, since the potential of eq 5 is used only to move the system to the pertinent region of configurational space. At the resulting geometries, the ground state energies E_g are correctly evaluated with the diagonalization procedure described previously, and the free energy changes are obtained with the umbrella-sampling approach described in the following.

The change in free energy associated with moving from one mapping potential to the next can be evaluated by the relationship[37,38]

$$\delta G_{n\rightarrow(n+1)} = -\beta^{-1} \ln\{\langle \exp[-(\epsilon_{n+1} - \epsilon_n)\beta]\rangle_n\} \quad (6)$$

where $\langle\cdots\rangle_n$ denotes an average over trajectories on mapping potential ϵ_n. Equation 6 represents an FEP procedure, in which the state with potential ϵ_n serves as a reference state and the change of the potential to ϵ_{n+1} is considered to be a small perturbation. In principle, this treatment is exact if the steps are sufficiently fine.

Once the $\delta G_{n\rightarrow(n+1)}$ values are known, the free energy of the adiabatic ground state at the transition state or, more generally, at any particular value of the reaction coordinate (x') can be obtained by the umbrella sampling (US) formula[20,32]

$$\Delta g(x') = \Delta G_m - \beta^{-1} \ln\langle \delta(x - x') \exp[-\beta(E_g(x) - \epsilon_m(x))]\rangle_{\epsilon_m} \quad (7)$$

where ϵ_m is a mapping potential that yields x values close to x'; and ΔG_m is the cumulative free energy change for moving from ϵ_1 to ϵ_m:

$$\Delta G_m = \sum_{n=1}^{m-1} \delta G_{n\rightarrow(n+1)} \quad (8)$$

The exponential term in eq 7 is the Boltzmann factor for moving from ϵ_m to E_g at a given position on the reaction coordinate (x), and the Dirac delta function ensures that only points where $x = x'$ contribute to the average. If the evolution of the mapping potential is sufficiently gradual, the free energy functionals $\Delta g_g(x')$ obtained for several different values of m overlap for a given range of x'. In this case, patching together the full set of $\Delta g_g(x')$ gives the complete free energy curve for the reaction. The standard free energy change for a reaction ($\Delta G_{R\rightarrow P}$ can be calculated by eq 7 if x' represents the most probable value of x in the product state. More generally,

$$\Delta G_{R\rightarrow P} = -\beta^{-1} \ln\{\int_{x^{\ddagger}}^{\infty} \exp[-\Delta g(x)\beta]\, dx / \int_{-\infty}^{x^{\ddagger}} \exp[-\Delta g(x)\beta]\, dx\} \quad (9)$$

2.4. Parametrization of the EVB/FEP Calculations of the Activation and Reaction Free Energies. Our calculations examine the energetics of the associative nucleotidyl transfer from dGTP to the 3′-terminal deoxyribose of the primer DNA strand. These calculations involve comparing the energetics of the preequilibrium proton transfer from the 3′-OH nucleophile to three different acceptors. The acceptors considered include the ionized Asp654 residue, bulk water, and one of the oxygens bonded to the P_α atom. For the mechanisms with Asp654 serving as proton acceptor (general base), the subsequent nucleophilic attack and the departure of the leaving group are also evaluated. The definition of the reacting part (S in eq 4) for each of these mechanisms and the details of the EVB/FEP calculations are given in the following.

In general, the EVB parameters were judiciously chosen using numerous trial-and-error EVB calculations of the model reactions in gas phase and aqueous solution and the synthesis of information from the following sources: the parameters used in the previous EVB studies by Warshel[20] and Åqvist[10,11] groups, our unpublished gas-phase ab initio calculations of the proton transfer in the prototypical short-strong (also called low-barrier) hydrogen bonds, ab initio calculations on the hydrolysis of dimethyl phosphate (see section 3.2 and ref 41) and methanolysis of the phenyl phosphate dianion in solution,[42,43] scaled quantum mechanical force field of sodium dimethyl phosphate,[44] and ab initio calculations of the variations of atomic charges during proton-transfer reactions in hydrogen-bonded systems.[45] The complete list of the resulting EVB parameters is presented in Tables 1S−9S of the Supporting Information.

The FEP calculations of each individual reaction step were carried out for trajectories having total lengths of 125, 250, and 600 ps calculated with the time step of 1 fs. These trajectories were divided into 51 frames that differed in the value of λ driving the reaction from the initial to the final VB state (eq 5). Energies were collected every 10 fs, and the first 10 energies in each frame were discarded prior the free energy evaluation. Gas-phase shift and off-diagonal EVB elements were adjusted to reproduce the correct energetics of the solution reactions (see also the Results section). The methodological aspects of the configuration sampling by molecular dynamics are described in section 2.5.

The C3′−OH part of the 3′-terminal nucleotide of the primer, α- and β-phosphate groups of the dGTP substrate, and the COO^- group of Asp654 were selected as a reactive ("quantum") part (Figure 1). For these fragments, the geometry, connectivity,

(37) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 5, pp 169−194.
(38) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(39) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(40) Radmer, R. J.; Kollman, P. A. *J. Comput. Chem.* **1997**, *18*, 902−919.
(41) Jenkins, L. A.; Bashkin, J. K.; Pennock, J. D.; Florian, J.; Warshel, A. *Inorg. Chem.* **1999**, *38*, 3215−3222.
(42) Florián, J.; Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 11524−11525.
(43) Florián, J.; Warshel, A. *Phosphorus, Sulfur Silicon Relat. Elem.* **1999**, *146*, 525−528.
(44) Florián, J.; Baumruk, V.; Strajbl, M.; Bednarova, L.; Stepanek, J. *J. Phys. Chem.* **1996**, *100*, 1559−1568.
(45) Florián, J.; Scheiner, S. *J. Comput. Chem.* **1994**, *15*, 553−560.

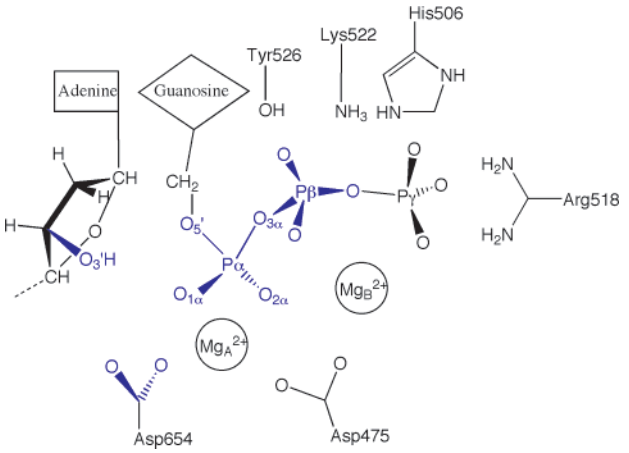

Figure 1. Schematic structure and numbering of the dGTP substrate and nearby active site residues. The reactive part that was allowed to change the bonding pattern and atomic charges in the simulated reactions is drawn in blue. Partial atomic charges, except those of Mg^{2+}, are not shown.

and atomic charges were allowed to vary among the different VB structures. Seven different VB states were considered (Figure 2): the reactant state (I), the intermediate containing protonated Asp654 and the negatively charged nucleophile (II), the intermediates in which both Asp654 and nucleophile are deprotonated (V and VI), the pentacoordinated phosphorane intermediates (III and VII), and the product state (IV) (Figure 2).

2.4.1. Mechanism with the Aspartate 654 Serving as the General Base. The reaction steps I → II, II → III, and III → IV, encompassing the complete general base-catalyzed nucleotidyl transfer reaction, were studied by three separate EVB/FEP-US calculations. The second and third steps of the reaction in the protein used the final atomic positions and velocities obtained in the calculation of the first and second steps, respectively. The reaction I → II in the protein was initiated from the crystal structure geometry equilibrated by molecular dynamics simulations as described in section 2.5. The reaction I → II in water was modeled by the proton transfer from methanol to aspartate. The reactive part included the atoms shown in Figure 1 without the atoms belonging to GTP. The EVB parameters used were a pertinent subset of the parameters given in Tables 1S−9S. The distance between the proton donor and acceptor atoms in the reactant state (I) in water was constrained by a flat-well harmonic potential, with the flat (unconstrained) region between 1.0 and 2.8 Å, and 2 kcal/mol harmonic force constant.

The reactions II → III and III → IV in water were calculated for the system consisting of CH_3O^-, aspartic acid residue, GTP^{4-}, and two Mg^{2+} ions. The reacting part (Figure 1) and all the EVB parameters (Tables 1S−8S) were identical to those used for the calculations in the protein. The distance between O_3' (nucleophile) and P_α atoms in water reactions was constrained by a flat-well harmonic potential, with the flat (unconstrained) region between 1.0 and 3.3 Å. The walls beyond this region were determined by the 10 kcal/mol force constant. Similarly, the distance between the P_α atom and the oxygen of the leaving group ($O_{3\alpha}$ in Figure 1) atoms was constrained by a 1.0 Å, 4.4 Å, 10 kcal/mol flat-well harmonic potential. These constraints limited the reactants to be in a common solvent cage during the entire simulation. The corresponding 2 kcal/mol cage corrections (eq 3) were added to the calculated activation barriers

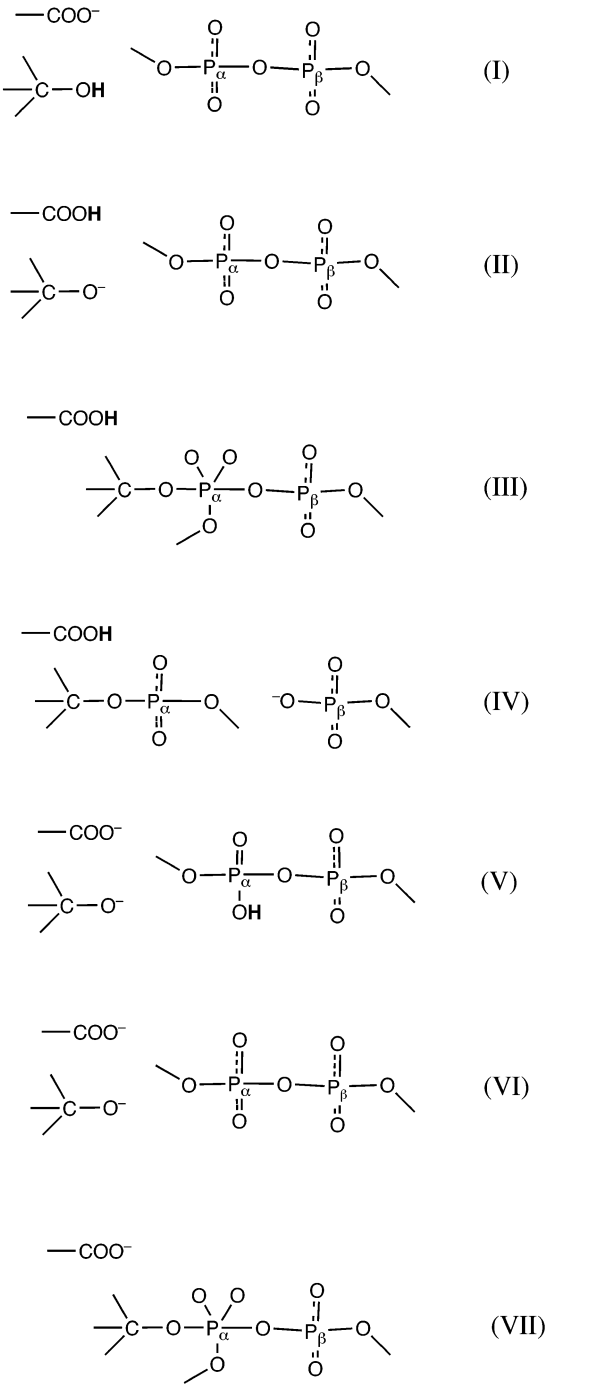

Figure 2. Valence bond structures representing reactants, intermediates, and products of the reactions.

in water. No distance or positional constraints were used in the protein calculations.

2.4.2. Mechanism with the Substrate Serving as the Base. The C3′−OH part of the terminal primer nucleotide and α-phosphate group (PO_4) were selected as the reactive part. For these fragments, the geometry, connectivity, and atomic charges were allowed to vary among different VB structures. Two VB states were considered. These states involved the reactant state (I) and the intermediate (V) containing the protonated $O_{1\alpha}$ atom of the phosphate and the deprotonated nucleophile (C3′−O3′$^-$) (Figures 2 and 1S). The $O_{1\alpha}$ oxygen was chosen as proton acceptor because it was closer to the nucleophile in the reactant state than the other nonbridging oxygen ($O_{2\alpha}$). The EVB/FEP-

US calculation of the reaction I → V in the protein was initiated from the crystal structure geometry equilibrated by molecular dynamics simulations as described in section 2.5.

The reaction in water was modeled by the proton transfer from methanol to α-phosphate of GTP, with the same reactive part as that for the protein reaction. The charges on the nonbridging oxygens of the β- and γ-phosphate were chosen in such a way that the β- and γ-phosphate groups were electroneutral. Thus, the overall charge of the solute was −1. The EVB parameters used were identical to those used in the protein calculation (Tables 1S−9S).

To rationalize our selection of using neutral rather than negatively charged β- and γ-phosphates, we would like to point out that the energetics of the studied proton-transfer reaction is determined by a pK_a difference between the 3′OH and α-phosphate groups (eq 11). In water, the pK_a of α-phosphate is known to be little affected by the protonation state of the β- and γ-phosphates or even their complete absence.[46,47] This experimental observation is difficult to reproduce by computer simulations because the simulation times and solvent sphere radius achievable by present computer technology are insufficient to provide a correct screening of high charges by the solvent. However, we have shown that in studies of solution reactions of highly charged systems we can get basically the same results by neutralizing distant charges and running relatively short calculations or by running long calculations with the full charge. Therefore, since we realize that the main parameters adjusted in solution are the local parameters of the Morse potential of the breaking and forming bond and the constant in the diagonal element of the EVB Hamiltonian (so-called gas-phase shift), we come to the conclusion that it is much more preferable to adjust these parameters by using stable calculations with neutralized charges on β- and γ-phosphates than using unstable simulations with full −1 and −2 phosphate charges.

The parameters obtained from the stable calculations are physically valid and thus fully transferable to the reaction in the protein active site. Of course, in the protein, we do retain the full charges. But in this case, the convergence is much faster than in the corresponding case in water, since the GTP charges are screened by nearby groups with opposite charges (e.g., the Mg^{2+} ions and positively charged protein residues). Thus, the neutralization of the charges of the β- and γ-phosphates in water should be simply viewed as a special procedure for an accelerated convergence and not as a selection of different parameters.

The distance between the proton donor and acceptor atoms in the reactant state (I) in water was constrained by a flat-well harmonic potential, with the flat (unconstrained) region between 1.0 and 3.6 Å, and 10 kcal/mol harmonic force constant. The position of the P_α atom was constrained (in the simulations of the solution reaction) to stay in the center of the simulation sphere.

2.4.3. Mechanism Involving Proton Transfer from the Nucleophile to Bulk Water. The free energy cost for deprotonating the neutral attacking group (3′-OH) in the protein active site and in aqueous solution was evaluated as described by Fothergill et al.[6] by using the relationships

$$\Delta G(\mathrm{ROH_{aq}} + \mathrm{OH^-_{aq}} \rightarrow \mathrm{RO^-_{aq}} + \mathrm{H_2O_{aq}}) = 2.303RT(\mathrm{p}K_a(\mathrm{ROH})_{aq} - \mathrm{pH}) \quad (10a)$$

$$\Delta G(\mathrm{ROH_p} + \mathrm{OH^-_{aq}} \rightarrow \mathrm{RO^-_p} + \mathrm{H_2O_{aq}}) = 2.303RT(\mathrm{p}K_a(\mathrm{ROH})_{aq} - \mathrm{pH}) + \Delta\Delta G_{sol}(\mathrm{ROH} \rightarrow \mathrm{RO^-})_p - \Delta\Delta G_{sol}(\mathrm{ROH} \rightarrow \mathrm{RO^-})_{aq} \quad (10b)$$

where aq and p indicate that the given system is immersed in bulk water and protein environments, respectively. The changes in solvation free energy, $\Delta\Delta G_{sol}(\mathrm{ROH} \rightarrow \mathrm{RO^-})$, in water and T7 pol were evaluated by FEP calculations of the free energy associated with a charging process, in which the charges of the C3′, O3′ and H3T atoms were changed from 0.0713, −0.6549, and 0.4396 au in the neutral state to −0.2891, −0.8549, and 0.0 au in the deprotonated state, respectively. All other parameters remained the same in both states. In the calculations of the deprotonation in aqueous solution, the substrate was modeled by 2′-deoxyribose and its H3′ atom was constrained in the center of the simulation sphere.

Although, in principle, one should consider the activation barrier in the proton transfer process (since this barrier might involve a sizable contribution from a penetration of OH^- to the protein and its reentry as H_2O to the bulk water). This barrier was not evaluated. However, as long as the proton transfer (PT) to the bulk water is not the rate limiting step of our reaction, it is sufficient to calculate only the reaction free energy of the PT step.

The reaction VI → VII (nucleophilic attack step) was studied using the same reacting part, topology, and EVB parameters as reaction II → III described in section 2.4.1 with the following exceptions. In all states, charges of the H3T, OD1, and OD2 atoms (Figure 1S) were set to 0.0, −0.8, and −0.8 au, respectively. The charge of the CG atom was 0.8 and 0.72 in the states VI and VII, respectively. The R* and ϵ vdW parameters of the H3T atom were set to zero, and all the angles and torsions involving this atom were removed from the topology. The OD1 atom was assigned the O2 atom type in both VI and VII states. The starting geometry for the calculation of the VI → VII reaction in the protein was obtained by 150 ps relaxation of the structure of the protein/DNA/dGTP complex in state II using the charges and EVB parameters of the state VI.

2.5. Molecular Dynamics Simulations. The configurational ensembles for the evaluation of the reaction energetics as described in section 2.1 were generated from the molecular dynamics (MD) trajectories using the AMBER force field[48] implemented in the program Q, version 4.27.[49] The water molecules were approximated by the TIP3P potential.[50] Starting structures for MD simulations were prepared from the crystal structure of the T7 DNA polymerase complexed with the DNA and ddGTP (1T7P),[51] as described in detail in our previous papers.[4,52]

(46) Serjeant, E. P.; Dempsey, B. *Ionization constants of organic acids in aqueous solution*; Pergamon Press: Oxford, 1979.
(47) Åqvist, J.; Kolmodin, K.; Florian, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71−R80.
(48) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(49) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J. Mol. Graphics and Modeling* **1999**, *16*, 213−225.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(51) Doublie, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. *Nature* **1998**, *391*, 252−258.

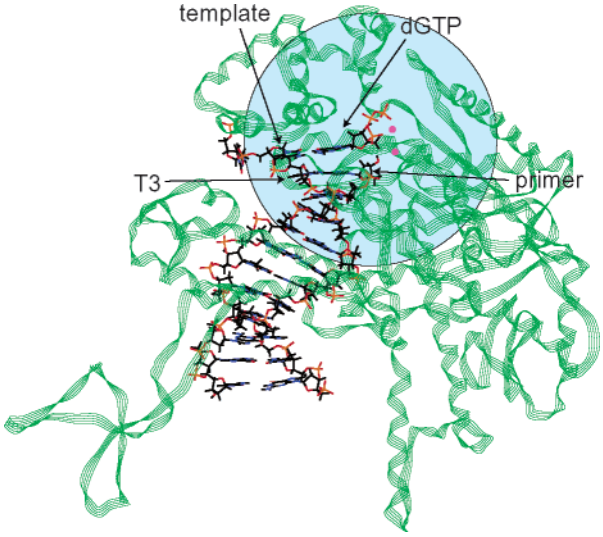

Figure 3. Illustration of the proportions of the simulated sphere with respect to the size and shape of T7 polymerase (green) and its primer-template DNA and dGTP substrates (black). The simulated region was filled with water molecules (blue).

In essence, this initial structural manipulation involved adding 3′-OH groups on the deoxyribose moieties of the ddGTP substrate and the terminal primer dideoxynucleoside, ionizing selected protein residues[4] and calculating an unconstrained 1.8 ns MD trajectory for all atoms in the 18 Å simulation sphere centered on the N1 atom of dGTP. This trajectory coincides with the production trajectory of ref 4. At this point, the center of the simulation sphere was moved to coincide with the P_α atom of the dGTP substrate, and the radius of the simulation sphere was increased to 21 Å.

New water molecules were added into the simulated system by immersing the simulation sphere into the sphere of bulk water molecules and removing those water molecules that were sterically overlapping with the atoms already present in the simulated system. The coordinates of the protein and DNA atoms protruding beyond the 21 Å sphere were fixed at their initial values. The water molecules were treated by the surface constraint all-atom solvent (SCAAS) model.[53] The nonbonded interactions of the atoms outside the simulation sphere with the atoms inside the simulation sphere were turned off. The total charge inside the simulation sphere was set to zero by neutralizing negative charges of three phosphate groups that lie very close to the edge of the simulation sphere.

To explain the reasons that led us to this uncharging, we would like to point out that negative phosphates of the DNA backbone are typically balanced by the nearby positively charged protein residues. However, in this case, the positively charged groups remained outside of the simulation sphere, and thus, the net negative charges contributed by these phosphate groups were actually artifacts of the introduction of the spherical boundary. The position and the size of the simulated sphere with respect to the protein size are illustrated in Figure 3.

At this stage of the refinement of the structure of the protein/DNA/substrate complex, we also changed the charges and vdW radii of the Mg^{2+} atoms and ionic oxygens on phosphate and carboxyl groups (with atom types O2 and O3). This change was motivated by the fact that, during a 1.6 ns equilibration, the average calculated Mg−O distances in the first coordination sphere were too short (i.e., about 1.8 Å calculated versus 2.1−2.3 Å observed in the crystal). To fix this discrepancy, we increased the vdW radius (R* parameter) of the oxygens having the O2 type (in COO− groups) from 1.66 to 1.90 Å, those having O3 type (in $-PO_2^--$ or $-PO_3^{2-}$ groups) from 1.70 to 1.90 Å, and R* of Mg^{2+} ion from 1.1 to 1.3 Å. The latter change had to be compensated by increasing the charge on the magnesium atoms from 1.66 to 2.0 au in order to keep the solvation free energy of Mg^{2+} close to its experimental values. The solvation free energy of Mg^{2+}calculated with the new charge and R* parameters was −425 kcal/mol. This energy is in reasonable agreement with the experimental estimate of −454 kcal/mol (see ref 29 and references therein).

The solvation free energy of acetate calculated with the oxygen vdW radius of 1.9 Å was −85 kcal/mol, which is also reasonably close to the experimental value of -82 ± 5 kcal/mol.[28] The solvation free energy calculations were carried out using the FEP method with an overall simulation length of 30 ps. To obtain total solvation free energy, the contribution from the solvent molecules lying beyond the 21 Å simulation sphere was estimated using the Born formula and added to the calculated FEP results. Larger positive charges of magnesium atoms were compensated by making charges of the seven ionic oxygens of dGTP more negative. The charges of the triphosphate moiety of dGTP are given in Table 1S. The total charge of dGTP + $2Mg^{2+}$ was zero.

After making parameter modifications described above, the equilibration of the protein/DNA/water system was continued in the 21 Å simulation sphere using the following protocol. First, the solvent and the substrate were relaxed by a series of three short-step (0.05−0.5 fs) unconstrained simulations at 10 K (total 0.6 ps simulation time). These calculations were followed by a gradual heating of the simulated system from 10 K to 298 K in a series of 10 unconstrained MD simulations with a 1 fs step and an 85 ps total simulation time. The equilibration protocol was concluded with a 100 ps trajectory calculated at 298 K. The "shake" algorithm was applied to hydrogen atoms of both solute and solvent molecules that did not belong to the reacting part. All nonbonded interactions of the reacting part were evaluated explicitly (no cutoff). Other nonbonded interactions were evaluated explicitly for distances shorter than 10 Å. The local-reaction field (LRF) method[54,55] was used to include long-range electrostatic interactions for distances beyond the 10 Å cutoff.

3. Results

3.1. Ground State Structure of the Catalytic Site. The substrate and 3′-primer terminal nucleotide present in the crystal structure of the T7 pol/DNA/ddGTP complex do not have the same chemical structure as those present in native polymerase since they lack the 3′-OH groups on the sugar moieties. Crystallizing T7 polymerase with 2′,3′-dideoxy sugars helps to obtain high resolution structures by curtailing enzymatic activity in the crystal but hinders significantly our ability to assess the catalytic mechanism directly from the observed crystal structure.

(52) Florián, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 5739−5753.
(53) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124−129.
(54) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100−3107.
(55) Sham, Y. Y.; Warshel, A. *J. Chem. Phys.* **1998**, *109*, 7940−7944.

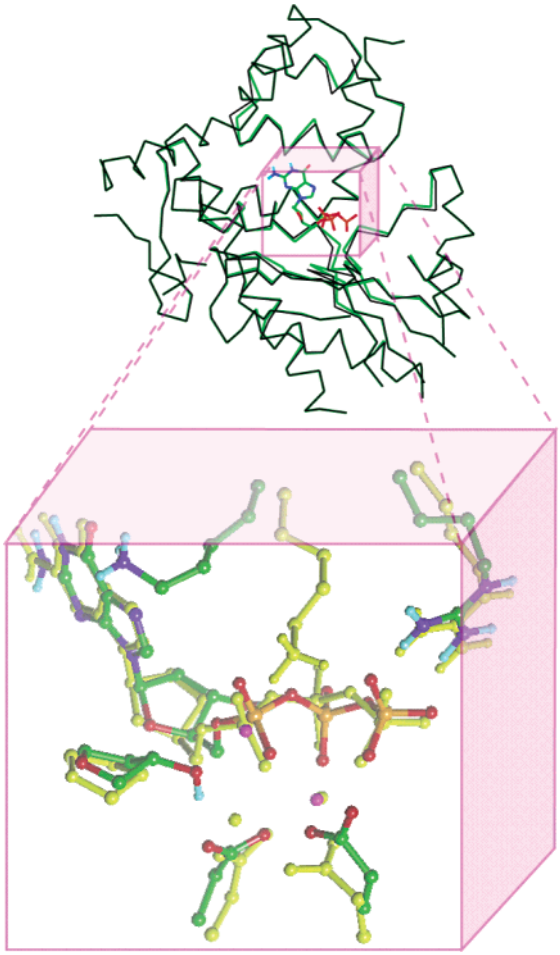

Figure 4. (Upper part) Comparison of the X-ray (black) and calculated (green) ground state structures of the backbone C_α atoms of T7 polymerase. (Only atoms that lie within 20Å from P_α are shown.) (Lower part) Comparison of the X-ray (yellow) and calculated (atom based colors) structures of the active site of T7 pol with bound dGTP substrate.

This structural "imperfection" is eliminated in our ground state computer simulations as well as in our simulations of reaction mechanisms reported in section 3.3.

The ground state geometry of the active site calculated with the correct dGTP substrate and primer sugars is depicted in Figure 4. As can be seen from the upper part of Figure 4, the observed protein backbone conformation is retained after 1.8 ns of unconstrained MD simulation. However, substantial deviations from the X-ray structure occur for the active site region (Figure 4, lower part). Here, the largest changes involve average positions of the Mg_A^{2+} ion and the side chain of Lys522 (see Figure 1 for residue numbering). Both structural changes are clearly induced by the presence of the 3′-OH group of the substrate, which interacts unfavorably with the Mg_A^{2+} ion in its crystallographic position.

The shift of the Mg_A^{2+} ion to a new position brings it closer not only to the second nonbridging oxygen of the α-phosphate but also to the NH_3^+ group of Lys522. The resulting electrostatic repulsion between the positively charged Mg_A^{2+} and NH_3^+ groups is relieved by rotating the Lys522 side chain around the $C_\alpha-C_\beta$, $C_\gamma-C_\delta$, and $C_\delta-C_\epsilon$ bonds. This conformational change occurs in the 500 ps−550 ps time frame of the unconstrained equilibration trajectory, and the new position of Lys522 is retained in the subsequent part of the equilibration trajectory and also during the whole EVB calculation. The loss of a favorable electrostatic interaction between the NH_3^+ group and the nonbridging α-phosphate oxygen is compensated by interactions between the NH_3^+ group and the N7 atom of the guanine moiety of dGTP. Although the ligands of the Mg_A^{2+} ion change, for example, Asp654 no longer coordinates Mg_A^{2+}, the octahedral character of the Mg_A^{2+} coordination sphere is retained in the calculated structure. The new ligands and their average distances from Mg_A^{2+} are the O3′ atom of the primer (2.2 Å), two nonbridging oxygens of the α-phosphate (2.2 and 2.3 Å), and oxygens of three water molecules (2.1 Å; these water molecules are not shown in Figure 4 to avoid overcrowding).

The O3′ oxygen is positioned 3.3 Å from the P_α atom, and thus, it is ready for the nucleophilic attack of α-phosphate. However, at neutral solution pH, O3′ carries a hydrogen that hinders the formation of a new $O3'-P_\alpha$ bond. Thus, the nucleophilic attack must be preceded or accompanied by the proton transfer from the O3′H group to a suitable base. The calculated ground state structure shows the presence of a hydrogen bond between this group and an oxygen atom of the Asp654 residue. This ground state feature suggests that Asp654 might serve as a general base in the reaction catalyzed by the T7 polymerase.

However, another viable mechanism could involve transfer of the proton from the 3′OH group to the OH^- ion solvated by bulk water (see section 2.4.3). One of the nonbridging oxygens of the α-phosphate could also serve as a proton acceptor. Our selection among three possible mechanisms is based on calculations of the appropriate activation free energies in water (section 3.2) and in the enzyme active site (section 3.3).

3.2. Energetics of the Reference Reactions in Water. A central aspect of our enzyme simulations is the comparison of the free energy profiles of enzymatic reactions with the corresponding reference reactions in water using the same computational methodology,[20,56] thus enabling a quantitative estimate for the free energy of the nucleotide insertion reaction.

This approach has conceptual as well as practical advantages. Conceptually, it allows a "best" estimate to be made for the enormous catalytic power of enzymes[57] and provides a rigorous approach to evaluate alternative models describing the source of this catalytic power.[25,58−60] The practical advantage stems from the fact that none of the current computational methods are capable of yielding energies for biologically relevant processes with a sufficient absolute accuracy. Thus, the error cancellation between identical reactions calculated for different environments becomes essential for obtaining quantitatively correct free energies for enzymatic reactions. In the EVB approach, key parameters H_{12} and α_0 are adjusted so that the energetics of the reference reaction in water is correctly reproduced, and these parameters (along with other force field parameters) are then used without any modifications in the calculation of the enzyme energetics as described in section 2.

3.2.1. Proton-Transfer Reactions. All three proton-transfer reactions considered here (I → II, I → V, and I → VI) are nonspontaneous processes in aqueous solutions. Their experi-

(56) Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5250.
(57) Wolfenden, R.; Snider, R. J. *Acc. Chem. Res.* **2001**, *34*, 938−945.
(58) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035−27038.
(59) Warshel, A.; Florián, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5950−5955.
(60) Cannon, W. R.; Benkovic, S. J. *J. Biol. Chem.* **1998**, *273*, 26257−26260.

Table 1. Activation and Reaction Free Energies and EVB Parameters for Reference Reactions in Aqueous Solution

reaction[a]	type[b]	ΔG_0 (kcal/mol)	$\Delta g^{\ddagger}$ (kcal/mol)	H_{12}[c] (kcal/mol)	α_0[c] (kcal/mol)
I → II	PT	12.6[d]	14.4	15.0	53.0
I → V	PT	16.0[e]	17.7	16.5	51.5
I → VI	PT	8.2[f]			
II → III[g]	NucA	18.5	19.4	37.0	195.9
III → IV	DepLg	−24.9	2.6	45.0	−287.4
VI → VII[g]	NucA	18.5	19.4	26.5	175.4

[a] Notation according to Figure 2. [b] Proton transfer (PT), O3′ attack on P_α (NucA), departure of the leaving group (DepLg). [c] H_{12} denotes the off-diagonal element of the EVB Hamiltonian. α_0 denotes the gas-phase shift of the product state, which appears in the H_{22} element of the EVB Hamiltonian (see eq 4). Note that the gas-phase shift in the reactant state (H_{11}) is always zero. The magnitudes of ΔG_0 and $\Delta g^{\ddagger}$ calculated using H_{12} and α_0 constants reproduce the observed (or empirically estimated) ΔG_0 and $\Delta g^{\ddagger}$ values presented in this table with the accuracy of 0.1 kcal/mol. [d] pK_a (donor) = 13.0;[98] pK_a (acceptor) = 3.8.[46] [e] pK_a (donor) = 13.0;[98] pK_a (acceptor) = 1.3.[99] [f] Experimental reaction free energy determined for pK_a (donor) = 13.0[98] and solution pH 7.0 using eq 10a. [g] The energetics of the solution reactions II → III and VI → VII are assumed to be identical.

mental *reaction free energy* in water is determined by the difference in the pK_a of the donor and the pK_a of the conjugate acid of the acceptor groups as

$$\Delta G_0 = 2.303RT(\mathrm{p}K_a\,(\mathrm{donor}) - \mathrm{p}K_a\,(\mathrm{acceptor})) \quad (11)$$

The accurate reproduction of the these ΔG_0 values by EVB simulations is a relatively straightforward task that requires adjusting the α_0 constant (gas-phase shift), which appears in the H_{22} term of the 2 × 2 EVB Hamiltonian (Table 1). On the other hand, actual *activation barriers* of the proton-transfer reactions in aqueous solution are more difficult to determine, both theoretically and experimentally.

Gas-phase ab initio calculations typically yield only single-minimum potential energy surfaces for proton transfers between oxygen atoms with ΔG_0 larger than 10 kcal/mol.[61] However, such calculations neglect the contribution of the solvent, which usually tends to stabilize the charge distribution of the reactant and product states more than that of the transition states (TSs) because TSs have more delocalized charge distributions.[62]

Additional non-negligible contributions to activation barriers of proton-transfer processes may originate from the nonequilibrium solvation effects.[63] Still, the activation barriers due to the solvent effects have to be very small since the rates of proton-transfer processes between oxygen atoms in water are limited only by diffusion.[64] For example, the rate constant observed for the proton transfer from the protonated ATP ($HATP^{3-}$) to OH^- is 1×10^9 M^{-1} s^{-1}.[64] Therefore, it is reasonable to position the activation barriers for the reactions I → II and I → V in water about 4 kcal/mol above the free energy of their product states. This barrier is reduced to about 2 kcal/mol when the reaction in the solvent cage ($\Delta g^{\ddagger}_{\mathrm{cage}}$, eq 3) is considered. The calculated $\Delta g^{\ddagger}_{\mathrm{cage}}$ values were set to agree with this value by adjusting the H_{12} element of the EVB Hamiltonian (Table 1). The calculated diabatic energies and the free energy profile for the proton-transfer reactions I → II and I → V are given in Figure 2S.

3.2.2. Formation and Cleavage of PO Bonds. Kinetic studies of model chemical reactions in water that constitute reference reactions for enzymatic phosphoryl and nucleotidyl transfer reactions are very difficult. This is because these reactions are typically very slow at room temperature[65−67] and, in many cases, their mechanistic interpretation is unclear.[47] Aside from trivial preequilibrium proton-transfer steps considered in the previous section, a fundamental reference reaction for the catalysis by DNA polymerases can be expressed as

$$\mathrm{dRib}^- + \mathrm{CH_3O{-}P_\alpha(O_2){-}O{-}P_\beta(O_2){-}O{-}P_\gamma O_3}^{4-} \underset{}{\overset{\mathrm{Mg^{2+}}}{\rightleftharpoons}} \mathrm{dRib{-}O{-}P(O_2){-}OCH_3}^- + \mathrm{P_2O_7}^{4-} \quad (12)$$

In this reaction, the P_α atom of methyl triphosphate ($CH_3P_3O_{10}^{4-}$) forms a new bond with the 3′-oxygen of deoxyribose in its deprotonated state (dRib−), and the P_α−O bond in the P_α−O−P_β bridge is broken, yielding the phosphate diester and pyrophosphate ($P_2O_7^{4-}$) as products.

Unfortunately, the kinetics of this reaction have not been experimentally observed, most probably because the attack on the α-phosphate is significantly slower than the attack on the γ-phosphate. To avoid this practical complication, Herschlag and Jencks[68] replaced the methyl triphosphate substrate by methyl 2,4-dinitrophenyl phosphate (MDP). The rate constant $k = 2.8 \times 10^{-4}$ M^{-1} s^{-1} observed by Herschlag and Jencks for the reaction of OH^- with MDP at 39 °C[68] can be converted into the activation free energy of a rate-limiting step ($\Delta g^{\ddagger}_{\mathrm{cage}}$) using TST (eq 1).

Application of TST to the OH^- attack on MDP yields $\Delta g^{\ddagger}_{\mathrm{cage}} = 21.1$ kcal/mol (eqs 1 and 3). To transform this barrier into the barrier for the reaction 12 (see eq 12), we assume that the hydroxide and deoxyribose anions have similar nucleophilicities (when they both are at 1 M concentration). We must also consider corrections for the difference in the pK_a's of the leaving group of the 2,4-nitrophenol (pK_a 4) and pyrophosphate (pK_a 8.4) and for the effect of magnesium counterions. Here, the addition of Mg^{2+} ions is known to accelerate reactions of phosphate diesters in aqueous solution by 1 to 3 orders of magnitude,[69] but the higher pK_a of pyrophosphate should slow the reaction by approximately the same amount.[70] Thus, we assume that these two effects will approximately cancel each other. This estimate leads to the 21.1 kcal/mol activation free energy of the rate-limiting step for reaction 12.

After estimating the overall activation barrier for reaction 12 in water, it is necessary to clarify the structure of the corresponding transition state and other mechanistic details. Reaction 12 involves the formation of the new chemical bond between the O3′ oxygen of deoxyribose and P_α phosphorus of methyl triphosphate, as well as the cleavage of the P_α−O bond in the P_α−O−P_β linkage. This bonding rearrangement can be accomplished in two ways. The scenario, in which the bridging P_α−O bond is broken in the first step of the reaction, is a dissociative mechanism. Although the dissociative mechanism

(61) Florián, J. Unpublished results.
(62) Scheiner, S.; Kar, T. *J. Am. Chem. Soc.* **1995**, *117*, 6970−6975.
(63) Villa, J.; Warshel, A. *J. Phys. Chem.* **2001**, *105*, 7887−7907.
(64) Eigen, M. *Angew. Chem., Int. Ed. Engl.* **1964**, *3*, 1−72.
(65) Bunton, C. A.; Mhala, M. M.; Oldham, K. G.; Vernon, C. A. *J. Chem. Soc.* **1960**, 3293−3301.
(66) Wolfenden, R.; Ridgway, C.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 833−834.
(67) Williams, N. H.; Wyman, P. *Chem. Commun.* **2001**, 1268−1269.
(68) Herschlag, D.; Piccirilli, J. A.; Cech, T. R. *Biochemistry* **1991**, *30*, 4844−4854.
(69) Blasko, A.; Bruice, T. C. *Acc. Chem. Res.* **1999**, *32*, 475−484.
(70) Kirby, A. J.; Younas, M. *J. Chem. Soc. B* **1970**, 510−513.

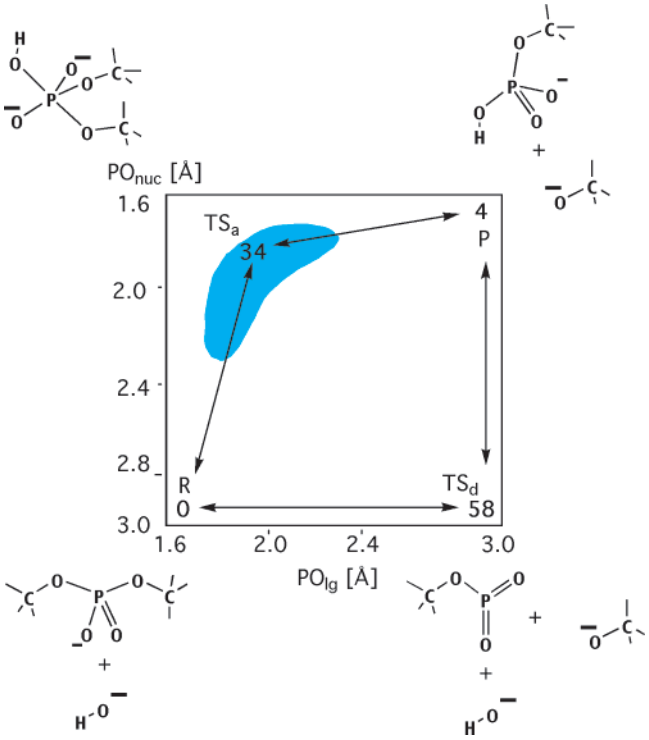

Figure 5. Free energy profile (Δg_{cage}, kcal/mol) for the hydrolysis of dimethyl phosphate at pH 15.6 (1 M OH^-) calculated at the MP2/6-31G**//HF/6-31G*+LD level of theory. The blue area around TS_a indicates a flat surface with free energies less than 1.5 kcal/mol below TS_a.

is not uncommon in phosphorus chemistry involving monoester types of substrates.[31,71−73] it appears to be less probable for reactions of substrates in the diester class, for example, the P_α phosphate group of dNTP.[70]

In the second (associative) mechanism, the nucleophilic attack occurs prior to the departure of the leaving group and an unstable pentavalent phosphorus species (phosphorane) appears on the reaction coordinate. The phosphorane species may correspond either to a high-energy intermediate or a transition state.

Since the detailed shape of the free energy profile is exceedingly difficult to determine unambiguously from the kinetic experiments, we carried out a series of ab initio calculations for the hydrolysis of the dimethyl phosphate anion by the hydroxide ion. Approximation of the pyrophosphate leaving group by methanol was necessary to keep the overall size of the system in proportions manageable at the correlated ab initio theoretical level. The resulting free energy profiles along the associative (TS_a) and dissociative (TS_d) reaction pathways connecting the reactant (lower-left corner) and product (upper-right corner) states are presented in Figure 5. We found that, for alkaline hydrolysis in aqueous solution, TS_d has a 24 kcal/mol higher energy than TS_a, and thus, the pure dissociative pathway can be excluded with a high degree of "certainty". This result is also valid for the neutral hydrolysis of dimethyl phosphate (pH 7), where the associative pathway is preferred by 12 kcal/mol.[61] The associative pathway is also favored by 23.5 kcal/mol in the active site of pol β modeled by several active site residues.[17]

The calculated activation free energy of 34 kcal/mol for the associative mechanism (Figure 5) in water, agrees well with the barrier of 32.4 kcal/mol corresponding to the rate constant of 6×10^{-7} M^{-1} s^{-1} at 125 °C observed for the alkaline hydrolysis of dimethyl phosphate by Westheimer et al.[74,75] However, it is difficult to determine with certainty, either experimentally (ref 71 and references therein) or by ab initio calculations, whether the actual associative mechanism of reaction 12 and related diester hydrolysis reactions is concerted as in Figure 5 or stepwise. The gas-phase ab initio calculations located dianionic phosphorane species as local minima on the potential energy surface,[76,77] whereas our ab initio calculations in aqueous solution did not lead to a stable dianionic phosphorane intermediate (Figure 5).

However, in a structurally analogous case of the hydrolysis of phosphate monoester dianions, we found that a concerted associative reaction turned into a stepwise reaction once the methanol leaving group was replaced by phenol.[31,42] Note that the calculated "dip" in the free energy surface corresponding to the phosphorane intermediate was so small that the surface would still appear to be concerted in many kinetic experiments.

Thus, given the accuracy limits of the ab initio calculations for reactions in solutions and ambiguities in mechanistic interpretation of kinetic experiments,[11] currently the only certainty about the nature of the dianionic phosphorane species is that, regardless of the leaving group, the reaction surface of the mid region of the associative pathway is essentially flat. The blue area in Figure 5 indicates this finding. Considering the large observed correlation between the pK_a of the leaving group and the logarithm of the rate constant for the hydrolysis of diaryl phosphates (Brønsted $\beta_{lg} = -0.67$ to -1.38),[70,78] we find it likely that the rate-limiting step for reaction 12 is dominated by the cleavage of the P−O bond to the leaving group.

In summary, our EVB calculations of the PO bond formation and cleavage reactions assume that the free energy profile for reaction 12 in water has the following features:

(i) The reaction proceeds via an associative mechanism.

(ii) The phosphorane region of the surface is essentially flat.

(iii) The rate-limiting step is shifted toward the departure of the leaving group.

(iv) $\Delta g^{\ddagger}_{cage}$ of the rate-limiting step is 21.1 kcal/mol.

(v) The overall reaction free energy is −6.3 kcal/mol (given by the difference in the stability of the nucleophile and leaving group in their deprotonated states).

These properties of the reaction surface are in a good accord with the actual calculated EVB surfaces (Table 1).

3.3. Reaction Energetics in the T7 Pol Active Site. The calculated energetics of several reaction steps, including a comparison of the results obtained for different lengths of MD trajectories, are presented in Table 2. The proton transfer (PT) from the 3′-OH group of the primer sugar to Asp654 is associated with the activation barrier ($\Delta g^{\ddagger}$) of 11 kcal/mol and

(71) Hengge, A. C.; Tobin, A. E.; Cleland, W. W. *J. Am. Chem. Soc.* **1995**, *117*, 5919−5926.
(72) Herschlag, D.; Jencks, W. P. *J. Am. Chem. Soc.* **1987**, *109*, 4665−4674.
(73) Hu, C.-H.; Brinck, T. *J. Phys. Chem. A* **1999**, *103*, 5379−5386.
(74) Haake, P. C.; Westheimer, F. H. *J. Am. Chem. Soc.* **1961**, *83*, 1102−1109.
(75) Kumamoto, J.; Cox, J. R., Jr.; Westheimer, F. H. *J. Am. Chem. Soc.* **1956**, *78*, 4858.
(76) Dejaegere, A.; Lim, C.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 4353−4355.
(77) Uchimaru, T.; Tanabe, K.; Nishikawa, S.; Taira, K. *J. Am. Chem. Soc.* **1991**, *113*, 4351−4353.
(78) Williams, N. H.; Cheung, W.; Chin, J. *J. Am. Chem. Soc.* **1998**, *120*, 8079−8087.

Table 2. Calculated Activation and Reaction Free Energies for Individual Steps for the Different Reaction Mechanisms of T7 Pol

reaction step[a]	reaction type[b]	reaction mechanism[c]	ΔG_0 (kcal/mol)	$\Delta g^{\ddagger}$ (kcal/mol)	simulation time (ps)
I → II	PT	A	6.1	10.7	125
			5.5	10.8	250
			6.0	11.0	600
I → V	PT	C	30.6	32.8	125
			28.5	31.1	250
			23.3	25.5	600
			22.1	24.7	1000
I → VI	PT	B	7.7[d]		125
			8.1		250
			7.5		600
VI→ VII	NucA	B	10.4	13.2	125
			9.5	12.7	250
			8.6	12.5	600
II → III	NucA	A	2.5	5.8	250
			2.6	5.8	600
III → IV	DepLg	A	−7.2	3.6	200
			−7.6	3.7	600

[a] Notation according to Figure 2. [b] Proton transfer (PT), O3′ attack on P_α (NucA), departure of the leaving group (DepLg). [c] A, B, and C denote associative mechanisms, which consist of the proton transfer to Asp654, bulk water, and α-phosphate, respectively, followed by the nucleophilic attack on α-phosphate and the departure of pyrophosphate. [d] $\Delta G_0 = \Delta G_0^{w} + \Delta\Delta G_{solv}^{pol} - \Delta\Delta G_{solv}^{w}$, where $\Delta\Delta G_{solv}^{pol}$ and $\Delta\Delta G_{solv}^{w}$ are free energies for charging the nucleophile from its neutral to the anionic (deprotonated) state in the protein ($\Delta G_{solv}^{pol} = -82.5$ kcal/mol in 600 ps FEP calculation) and in water ($\Delta G_{solv}^{w} = -81.8$ kcal/mol in 600 ps FEP calculation), respectively. See also eq 10.

the reaction free energy (ΔG_0) of 6 kcal/mol. The magnitudes of these energies are lowered in T7 pol with respect to their magnitudes in aqueous solution (cf. Table 1).

For PT to the α-phosphate, the calculated magnitude of $\Delta g^{\ddagger}$ of 25 kcal/mol is by 10 kcal/mol larger than $\Delta g^{\ddagger}$ observed for an overall rate-limiting step of nucleotidyl transfer in T7 pol. Thus, it is possible to exclude the mechanism involving PT to the substrate on the basis of results calculated for the PT step only. Subsequent steps in this mechanism would clearly further increase the overall barrier. Consequently, the calculations of the energetics of these steps were not carried out.

The reaction free energy for PT from the 3′-OH ribosyl group to OH^- in the bulk water was estimated by calculating the difference in solvation free energy of the 3′-O−ribosyl group in aqueous and protein environments (eq 10). This approach does not allow us to determine $\Delta g^{\ddagger}$. However, the calculated magnitude of ΔG_0 for this process constitutes a sufficient piece of information, provided the PT step is not the rate limiting step for the overall nucleotidyl transfer reaction. The reaction free energies for the first step of the PT to the bulk water at pH 7 in T7 pol and in aqueous solution were calculated to be similar (c.f. 7.5 kcal/mol versus 8.2 kcal/mol, Tables 2 and 1, respectively).

The nucleophilic attack of the 3′-O^- group of the deoxyribose on the P_α phosphorus atom was evaluated for two different protonation states of the Asp654 residue. The calculations with neutral Asp654 represent the second step in the Asp654 as a base mechanism, whereas calculations with ionized Asp654 correspond to the second step of the PT to bulk water mechanism. These nucleophilic attack steps are denoted II → III and VI → VII in Table 2 (see also Table 1 and Figure 2). The calculated energies indicate that the nucleophilic attack is more facile if Asp654 is in its neutral form. In this scenario, the reaction proceeds with a 5.8 kcal/mol activation barrier. On the other hand, the nucleophilic attack in the PT to bulk water mechanism is predicted to be associated with an activation barrier of 12.5 kcal/mol.

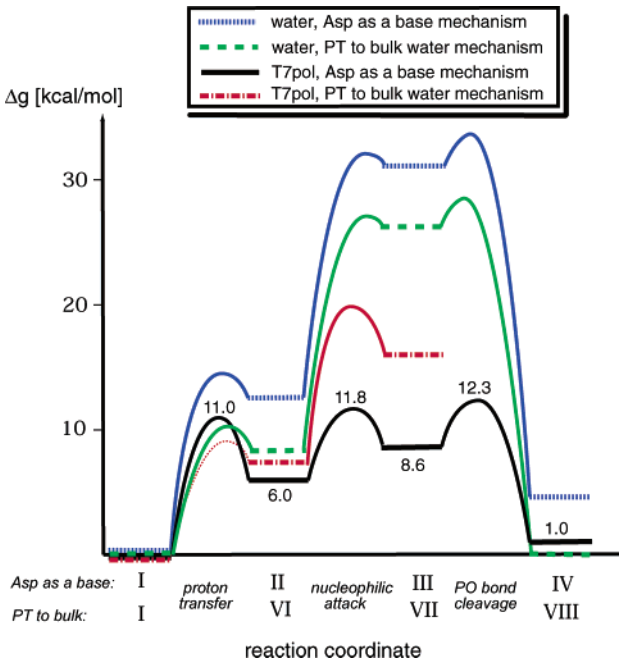

Figure 6. Comparison of the activation and reaction free energies along the reaction pathways for the "Asp as a base" mechanism in aqueous solution (blue) and T7 pol (black) and "PT to bulk water" mechanism in aqueous solution (green) and T7 pol (red). Note that the PT to bulk water mechanism favored in aqueous solution at pH 7 (green) is not favored in the protein active site. VB states on the reaction coordinate correspond to the states shown in Figure 2.

The departure of the leaving group from the pentavalent phosphorane intermediate in the Asp654 as a base mechanism (reaction III → IV) requires an additional 3.7 kcal/mol activation free energy. Given the large overall activation barrier for reaching the phosphorane intermediate in the PT to bulk water mechanisms (20.0 kcal/mol), the energetics for the departure of the leaving group in this mechanism was not examined. Clearly even in the most favorable case of a zero activation barrier for the leaving group departure, the PT to bulk water mechanism would be 7.7 kcal/mol less favorable than Asp654 as a base mechanism (Figure 6).

3.4. Transition State Structure for "Asp654 as a Base" Mechanism. The Asp654 as a base mechanism features three activation barriers that are almost isoenergetic (Figure 6). The highest of these barriers corresponds to the departure of the leaving group from the dianionic phosphorane intermediate. The average structure of the corresponding transition state is depicted in Figure 7. The bond lengths to the nucleophile and to the leaving group are 1.68 and 2.18 Å, respectively, indicating a late transition state with a dominant contribution of bond breaking. $O_{1\alpha}$, $O_{2\alpha}$, $O_{5'}$, and P_α atoms of the α-phosphate group retain a nearly perfect planar configuration characteristic of the pentavalent phosphorane intermediate. This arrangement is stabilized by a hydrogen bond to Asp654 and by the coordination of nonbridging oxygens with the Mg_A^{2+} ion, positioned 2.3 Å from both $O_{1\alpha}$ and $O_{2\alpha}$.

The second metal ion, Mg_B^{2+}, is coordinated to $O_{2\alpha}$ and to two oxygens of the pyrophosphate leaving group, with all three distances falling in the 2.2−2.3 Å range. These distances are

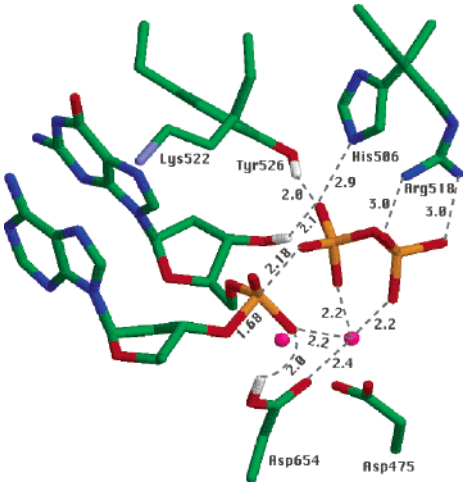

Figure 7. Transition state structure corresponding to the rate-limiting step of the nucleotidyl transfer reaction catalyzed by T7 pol. Selected interatomic distances (Å) are denoted by the dashed lines. H atoms bonded to oxygen are presented in gray color. H atoms bonded to the C and N atoms are not shown.

very similar to those observed in the crystal structure of the T7 pol ground state.[51] The leaving group is further stabilized by hydrogen bonding interactions with Arg518, Tyr526 and protonated His506 residues, and the 3′OH group of the dGTP substrate. All these hydrogen bonding interactions are already preformed in the ground state of T7 pol. Note that, although the interaction with the 3′OH group is missing in the crystal structure of T7 pol, which contains ddGTP substrate, it is present in our calculated ground state structure (Figure 4).

4. Discussion

The mechanism of the chemical step, which results in the addition of a nascent nucleotide at the 3′ end of a DNA primer strand, by DNA polymerases is known in the literature as a "two-metal-ion" mechanism.[79,80] This mechanism emphasizes the catalytic function of two Mg^{2+} ions present in the active site of all polymerases, while assigning the protein residues only a supporting role in the positioning of the catalytic metals.

Concerns about this mechanism have been voiced where it was suggested that one of the Asp residues in the active site of DNA polymerase β could actively participate in catalysis as proton acceptor.[14] However, a counterargument posits that the aspartic oxygen is too far removed from the 3′-OH group of the primer terminus and that the pK_a of the carboxyl group coordinated by the metal ion is inappropriate for a role as general base.[80] This latter view was then challenged[81] based on the argument that because the 3′-OH group is absent in the crystal structure,[14] the position of the aspartic side chain may be different when all the reactive groups are present in the active site. It was further stated that because both the donor and acceptor atoms in pol β are coordinated to Mg^{2+}, relatively small changes in metal−oxygen coordination geometry may be translated into large changes in pK_a values of the donor and acceptor groups.

Our stance on this key issue is that Mg_A^{2+} is indeed interacting with the 3′OH nucleophile but not with Asp654 in our calculated structure (Figures 4 and 3S), which presumably facilitates proton transfer to Asp654. Although the qualitative agreement between the previous intuitive predictions[14,81] and the ground state structure provided by our simulations is noteworthy, we realize of course that neither the ground state MD simulations nor the ground state crystal structures can provide conclusive information regarding the probability of reaching transition states via different mechanisms. Similarly, it is difficult to reach quantitative conclusions about the pK_a's in proteins based solely on structural information. In our view, it is essential to use some type of computer based structure−energy−function correlation to resolve these questions.

Nevertheless, establishing the correct ground state structure of the Michaelis−Menten complex has its own merit. Significant differences in the position of the Mg_A^{2+} ion and the side chain of Lys522 between the calculated structure and the X-ray structure[51] (Figures 4 and 3S) suggest that the crystal structure (and many crystal structures of ternary complexes of other polymerases with dideoxy-terminated primers) might not be biologically relevant. However, as pointed out by Doublie and Ellenberger,[82] the crystallographic position of Mg_A^{2+} ion agrees with the observed polymerase stereoselectivity for two diastereomers of α-phosphorothioate substrates (dNTPαS).[15]

In this regard, DNA polymerases incorporate only the *Sp* diastereomer of dNTPαS but not the *Rp* diastereomer. In the *Rp* diastereomer, the sulfur atom replaces the nonbridging oxygen coordinated by Mg_A^{2+} in the crystal structure.[15] (This oxygen is denoted as $O_{2\alpha}$ in Figure 1.) Since Mg_A^{2+} has a smaller affinity for sulfur than for oxygen,[15] the selectivity of DNA polymerases against *Rp* forms of dNTPαS substrates can be explained by less favorable substrate binding and/or by less favorable interactions of Mg_A^{2+} with the developing negative charges in the TS structure. In contrast, Mg_A^{2+} interacts nearly equally with both nonbriging oxygens of the P_α phosphate in the calculated structure (Figures 4 and 3S). Does this mean that the position of Mg_A^{2+} in the calculated structure contradicts the observed stereoselectivity for dNTPαS substrates? Not necessarily, because the stereoselectivity may be caused by interactions with the second metal ion, Mg_B^{2+}, which coordinates the $O_{2\alpha}$ oxygen but not $O_{1\alpha}$.

Stereoselectivity might also be caused by differences in conformational preferences of *Rp* and *Sp* isomers. For example, we found significant differences in the structure of *Rp* and *Sp* conformers of structurally related dimethyl ester of phosphorothioate anion in aqueous solution.[83] These conformational differences may result in large differences in the interactions of *Rp* and *Sp* isomers of dNTPαS with various active site amino acid residues.

It is noteworthy that our calculation can be used to explain the observed stereoselectivity by the presence of the hydrogen bond between the OH group of Asp654 and the $O_{2\alpha}$ nonbridging oxygen in the transition state structure (Figure 7). This hydrogen bond has to be replaced by a weaker OH···S interaction for the Rp-dNTPαS substrate. The strong hydrogen bond can also explain why the catalytic reaction for the Rp-dNTPαS substrate is not rescued by replacing Mg^{2+} counterions with Mn^{2+} ions, although Mn^{2+} ions show a higher affinity for sulfur.[15] In fact,

(79) Steitz, T. A. *Science* **1998**, *391*, 231−232.
(80) Steitz, T. A.; Smerdon, S. J.; Jager, J.; Joyce, C. M. *Science* **1994**, *266*, 2022−2025.
(81) Pelletier, H. *Science* **1994**, *266*, 2025−2026.
(82) Doublie, S.; Ellenberger, T. *Curr. Opin. Struct. Biol.* **1998**, *8*, 704−712.
(83) Florián, J.; Strajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959−7966.

Mn^{2+} ions do rescue the catalytic activity of the enzyme, but for the substrates substituted by sulfur on P_β rather than on P_α.[15] The observed data for dNTPβS substates can be explained in a straightforward manner by the interaction of $Mg^{2+}{}_B$ ion with nonbridging oxygens of P_β, as can be deduced from both the crystal structure and our TS structure (Figure 7). On the other hand, the observed kinetic data for dNTPαS substrates can be explained by the TS structure of Figure 7 but not by the X-ray structure.

Eventually, a crystal structure of the 3′OH-pol·DNA·dNTP$\alpha\beta$-(C) ternary complex, in which 3′OH denotes the 3′OH group on the terminal primer nucleotide and dNTP$\alpha\beta$(C) denotes a nonreactive dNTP analogue containing a methylene group in the $\alpha\beta$ bridge, might serve as an ultimate "judge" in the quest for the correct ground state structure. However, the Watson−Crick and mismatched template·dNTP base pairs may not have the same ground state structure.

In fact, some structural differences may be found even between different Watson−Crick base pairs, for example AT and CG, in the active site. Thus, we are at present uncertain about the geometry of the "correct" ground state structure or if any common ground state structure exists at all. However, we are confident that the structure presented in Figure 4 represents an important state on the free energy surface for the nucleotidyl transfer reaction in T7 pol. The importance of this state stems from the fact that it enables the Asp654 as a base reaction mechanism to proceed with a low activation barrier.

A shift in the position of $Mg_A{}^{2+}$ away from Asp654 is likely to represent a significant element of the mechanism of the polymerase selectivity even in situations when the free energy of this state is by a few kcal/mol higher than the crystal-like ground state. Future theoretical analysis will be required to examine the free energy difference and the activation barrier between the crystal-like and calculated-like structures (Figure 4).

For other enzymes, steady and presteady kinetic experiments are typically used to untangle mechanistic details of the chemical step of their catalytic reaction (see, for example, refs 13, 84−86). However, for DNA polymerases, the mechanistic interpretation of kinetic measurements is hampered by the presence of slow conformational changes flanking the chemical step. Thus, recent kinetic studies (see refs 87−89 and references therein) focus on revealing the relative rates of the chemical and conformational steps rather than on sorting out the details of the chemical step.

The interpretation of these experiments is controversial,[90] mainly because of the widespread use of small observed rate changes between natural and Sp isomers of dNTPαS substrates as means to exclude the slow chemical step. However, as shown by Herschlag et al.,[68] the rates of the nonenzymatic hydrolysis of the related oxygen- and sulfur-containing phosphate diesters are very similar. Consequently, there is no compelling reason for the assumption that their k_{cat} constants should be significantly different in DNA polymerases.

The importance of the protein conformational changes for the function of DNA polymerases has been illustrated by crystallographic studies[14,51] that showed, respectively, open and closed conformations for the DNA·polymerase complexes with and without cocrystallized correct dNTP substrate. However, at present, no crystal structures are available for catalytically competent ternary polymerase complexes with incorrect dNTP substrates. In our opinion, it is probable that the ground state Michaelis−Menten complex for the right and wrong dNTPs will involve significantly different protein conformations. Yet, we believe that even in this case the chemical step will be rate limiting for both the right and wrong dNTP substrates, for the following reasons. First, conformational changes typically have multiple "diffuse" pathways making it difficult for the protein to control their barriers. Second, DNA polymerases fall in several families that have vastly different amino acid sequences in the hinge region, but all of the polymerases have two conserved Asp residues and Mg^{2+} ions and identical templates in the active site. Thus, it is quite probable that these structural elements will play a major role in determining fidelity, whereas the other residues provide a supporting framework for this role. Third, there is a clear advantage in having a synergistic action of the dNTP substrate binding (K_D) and chemical (k_{cat}) contributions to fidelity, in analogy with the stepwise increase of polymerase fidelity resulting from nucleotide insertion and exonuclease proofreading.[91] The conformational change, if it was rate-limiting for the nucleotide insertion, would not benefit from such an incremental approach. Given the limitations encountered by experimental approaches mentioned above, a combination of biochemical and structural studies with theoretical free energy calculations seem at present as the most promising approach to address DNA polymerization and fidelity mechanisms.

The present work exploits the power of current computer simulation methods to convert structural information to activation free energies enabling the analysis of different mechanistic options in a quantitative way. The "Asp654 as a base" mechanism was found to be the only mechanism exhibiting "reasonable" energetics. This mechanism consists of three microscopic steps: (i) proton transfer to Asp654, (ii) nucleophilic attack of the C−O3′(−) group, and (iii) decomposition of the phosphorane intermediate by the departure of the pyrophosphate (PPi^{4-}). The feasibility of this mechanism relies on the local modification of the X-ray structure of the active site, which is needed to accommodate the 3′OH groups on the substrate and the primer nucleosides in a sterically and electrostatically favorable arrangement. This result should be understood in the following computational, biochemical, and structural context.

(1) The suggested mechanism was calculated for a Watson−Crick base pair between the incoming dGTP and the templating base (C). This base pair was present in the X-ray structure of the T7 pol ternary complex, which was used as a starting point for our MD calculations. Considering the possibility of significant structural differences between polymerases containing right and wrong dNTP substrates, as suggested by several groups,[88,92,93] we believe the actual reaction mechanism for wrong substrates

(84) Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5627−5631.
(85) Hunkapiller, M. W.; Forgac, M. D.; Richards, J. H. *Biochemistry* **1976**, *15*, 5581−5588.
(86) Stratton, J. R.; Pelton, J. G.; Kirsch, J. F. *Biochemistry* **2001**, *40*, 10411−10416.
(87) Carroll, S. S.; Benkovic, S. J. *Chem. Rev.* **1990**, *90*, 1291−1307.
(88) Johnson, K. A. *Annu. Rev. Biochem.* **1993**, *32*, 685−713.
(89) Arndt, J. W.; Gong, W.; Zhong, X.; Showalter, A. K.; Liu, J.; Dunlap, C. A.; Lin, Z.; Paxson, C.; Tsai, M.-D.; Chan, M. K. *Biochemistry* **2001**, *40*, 5368−5375.
(90) Showalter, A. K.; Tsai, M. D. *Biochemistry* **2002**, *41*, 10571−10576.
(91) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477−511.

might not involve Asp654 as a base. For example, if a significant misalignment of the dNTP in the active site occurs, deprotonation of the 3′-OH group by the OH^- ion from the bulk solvent could become more feasible than the general base mechanism. This is because the active site can be expected to become more water accessible as a consequence of base misalignment. Asp654 may shift away from the 3′-OH nucleophile, thereby raising the barrier for the Asp654 as a base mechanism. Reaction mechanisms for the wrong dNTP substrates are currently under investigation.

(2) The "substrate as a base" mechanism (mechanism C in Table 2), despite being used in several enzymes catalyzing GTP hydrolysis,[94] is disfavored in T7 pol. The proton transfer to the substrate is energetically disadvantageous in T7 pol probably due to the presence of two metals in the active site, as opposed to a single metal present in ATP- and GTP-ases. In contrast with our results, the substrate as a base mechanism was found to lead to a correct overall activation barrier of 16 kcal/mol in an ab initio quantum-mechanical theoretical study of the reaction mechanism of polymerase β by Abashkin et al.[17] The calculations of Abashkin et al.[17] modeled the enzyme as a gas phase cluster containing only the substrate and several amino acid side chains. This model, regardless of the sophisticated ab initio treatment, does not include important contributions made by the remainder of the environment (see, for example, refs 95 and 96) and is therefore unlikely to reveal the precise catalytic mechanism. The product energy of −19.1 kcal/mol (relative to the bound dNTP substrate) obtained by Abashkin et al. appears unreasonably low, making it highly unlikely that pol β could release the tightly bound product. Nevertheless, the active site geometries of T7 pol and pol β are sufficiently different and, thus, could in principle support different mechanisms for the chemical step of the nucleotidyl transfer reaction. EVB-based computer simulations are underway to address this possibility using X-ray structures obtained with ternary complexes of pol β.

(3) The computer simulations for T7 pol reveal that its catalytic amplification relative to water is about 15.3 kcal/mol. "Catalytic amplification" is defined as $\Delta g^{\ddagger}_{wat} - \Delta g^{\ddagger}_{cat}$, where $\Delta g^{\ddagger}_{wat}$ and $\Delta g^{\ddagger}_{cat}$ are the activation barriers for the reaction in water and in the enzyme, respectively.[58,59] The calculated catalytic amplification considers the change in mechanism upon going from water to the protein. In water, we suggest that the most probable mechanism involves PT to bulk water ($\Delta g^{\ddagger}_{wat} =$ 27.6 kcal/mol), whereas the "Asp654 as a base" is predicted to be the most likely mechanism in the active site of T7 pol ($\Delta g^{\ddagger}_{cat} =$ 12.3 kcal/mol). Because the actual observed rate of pol T7 turnover[97] corresponds to a $\Delta g^{\ddagger}_{cat}$ of about 15 kcal/mol (eq 1), our calculations overestimate the actual catalytic amplification of the enzyme by about 3 kcal/mol. This is an encouraging result considering the numerous sources of computational instability that must be dealt with.

An inspection of the average geometry of the TS suggests that the catalytic amplification of T7 pol originates primarily by having numerous protein residues in its ground state structure that can efficiently "solvate" the TS structure without having to undergo reorientation when leaving the ground state.

(4) Because the catalytic amplification for T7 pol compares the free energy *difference* between the reactions taking place in aqueous solution and the identical reactions occurring in the enzyme active site, it follows that the activation barriers reported for T7 pol are only as good (but possibly worse) as our estimates of the solution energetics. For the PT transfer steps, we estimate the uncertainty of the T7 pol energetics resulting from the uncertainty of the solution energetics as ±2 kcal/mol. A larger uncertainty of ±3 kcal/mol affects the energetics of steps involving PO bond formation or cleavage. This uncertainty may be eliminated when more accurate experimental or ab initio results for the model reaction in solution become available so that the predicted activation barriers in T7 pol can be readjusted without running additional EVB/FEP simulations.

Apart from calibration issues, there are several other factors that may diminish simulation accuracies. The most serious difficulty encountered by ours and all other current computer simulations of biomacromolecules emanates from instabilities in the calculated energies with respect to changes in simulation conditions. Critical simulation conditions include the number of sampled protein configurations, the boundary, the protonation state of the protein ionizable residues, and water penetration.

The calculated activation barriers for the three competing mechanisms studied in this work differ substantially. These significant activation barrier differences taken in conjunction with the satisfactory stability of the computational results suggest that the predicted "Asp654 as a base mechanism" is robust and therefore reliable.

(5) Free energy profiles encompassing multiple reaction steps with nearly equal activation barriers are often encountered in enzyme catalysis.[84] In accord with this "paradigm", three microscopic steps in the Asp as a base mechanism are predicted to have nearly identical activation barriers (Figure 6), imparting a degree of uncertainty to our designation of the departure of the leaving group as the rate-limiting step. Consequently, a rigorous computational determination of the source of polymerase fidelity should take into account substrate-dependent variations in the activation barriers for all three reaction steps depicted in Figure 6, rather than the rate-limiting step alone.

Alternatively, the simulations could focus on the stability of the pentavalent phosphorane intermediate, which is the structure that is most stabilized in the active site of T7 pol. Because this intermediate has a charge distribution similar to the TS for nucleophilic attack and leaving group release, a large part of this stabilization is still present at both TSs. This view is consistent with the TS structure presented in Figure 7. The TS structure retains the planar geometry of the nonbridging oxygen that is characteristic for this intermediate. There are, moreover, no H-bonding interactions of the protein residues with the leaving group oxygen. Thus, the polymerase strategy of catalyzing the nucleotide transfer reaction that emerges from

(92) Zhong, X. J.; Patel, S. S.; Werneburg, B. G.; Tsai, M. D. *Biochemistry* **1997**, *1997*, 11891−11900.
(93) Yang, L.; Beard, W. A.; Wilson, S. H.; Roux, B.; Broyde, S.; Schlick, T. *J. Mol. Biol.* **2002**, *321*, 459−478.
(94) Schweins, T.; Geyer, M.; Scheffzek, K.; Warshel, A.; Kalbitzer, H. R.; Wittinghofer, A. *Nat. Struct. Biol.* **1995**, *2*, 36−44.
(95) Bentzien, J.; Florián, J.; Glennon, T. M.; Warshel, A. In *Combined Quantum Mechanical & Molecular Mechanical Methods*; ACS Symposium Series 712; Gao, J., Thompson, M. A., Eds.; American Chemical Society: Washington DC, 1998; pp 16−34.
(96) Sicinska, D.; Paneth, P.; Truhlar, D. G. *J. Phys. Chem. B* **2002**, *106*, 2708−2713.
(97) Kumar, J. K.; Tabor, S.; Richardson, C. C. *J. Biol. Chem.* **2001**, *276*, 34905−34912.
(98) Izatt, R. M.; Rytting, J. H.; Hansen, L. D.; Christensen, J. J. *J. Am. Chem. Soc.* **1966**, *88*, 2641.
(99) Sigel, H.; Bianchi, E. M.; Corfu, N. A.; Kinjo, Y.; Tribolet, R.; Martin, R. B. *J. Chem. Soc., Perkin Trans. 2* **2001**, 507−511.

our calculations is that T7 pol stabilizes the phosphorane intermediate, which automatically brings down the free energy of the nearby transition states. This stabilization is applied up to the limiting point when the activation barriers for the PO bond formation and cleavage are similar to the activation barrier for the preceding proton-transfer step. The proton transfer to Asp654 is also catalyzed by the enzyme, but much less than the PO bond formation/cleavage steps. Thus, it is possible that, for certain genetically engineered pol T7 mutants, the proton-transfer step might become rate limiting. An observable deuterium kinetic isotope effect in the range of 1.5 to 3 could occur in just such a scenario.

(6) It is possible that the actual mechanism of PO bond formation/cleavage may be concerted. This means that there would be just a single TS on the free energy surface. Although the preceding point of this discussion, as well as our EVB model, assumes the existence of a pentavalent phosphorane intermediate, we do not expect the change of the mechanism from the stepwise to a concerted to change our overall conclusion. That is, "Asp654 as a base" would still be the mechanism for the PT step, and the pentavalent phosphorane structure would retain its importance for enzyme catalysis. The only difference brought about by a concerted mechanism would be the absence of a small dip on the free energy surface in the region of the dianionic pentavalent phosphorane. Regardless of whether the dianionic phosphorane is a TS or an intermediate, dianionic phosphorane is the structure that the enzyme has to stabilize to the largest extent.

We believe that the next major challenge will be to examine the effect of an incorrect dNTP substrate on k_{cat}. We recently studied the relative binding of right and wrong dNTP substrates in the active sites of T7 pol and Pol β.[4,52] If, as assumed in the present work, a chemical step is rate limiting for the overall nucleotide insertion reaction, then the sum of the calculated relative binding ($\Delta\Delta G_{bind}$) and chemical catalysis ($\Delta\Delta g^{\ddagger}$) steps will provide the total polymerase fidelity. Currently we believe that the presence of a bound incorrect dNTP destroys the preorganization in the active site. Although the active site can be restored to its correct preorganized structure, this will occur at the expense of destroying the binding site for the incorrect nucleotide and thus lead to unfavorable binding energy. This poor binding will result in an even higher total activation barrier (relative to the free energy level of the E + S system) than the one obtained from the active site with a destroyed preorganized structure. A cornerstone of any study designed to investigate polymerase fidelity should be a calculation of the relative stability of the dianionic phosphorane intermediate for the polymerase complexes containing right and wrong dNTP substrates. The activation energy of the proton-transfer step would need to be calculated, and if the "Asp654 as a base" mechanism becomes too slow for some mispairs, an alternative "PT to bulk water" mechanism might then take "center stage".

Acknowledgment. This work was supported by the NIH grants GM21422 (to M.F.G.) and GM24492 (to A.W.).

Supporting Information Available: Figures denoting atom numbering of the reacting part, diabatic and adiabatic free energy profiles for proton-transfer reactions I → II and I → V, and stereoscopic comparison of the X-ray and calculated ground state geometries of the T7 pol active site with bound dGTP substrate. EVB parameters for reacting fragments (9 tables). This material is available free of charge via the Internet at http://pubs.acs.org.

JA028997O