Supplemental Data

Polymerase β Simulations Reveal That Arg258 Rotation is a Slow Step Rather Than Large Subdomain Motions *Per Se*

Alternative title: **DNA Polymerase β Dynamics: Identification of Local, Rate-limiting Conformational Changes That Enhance Fidelity**

Linjing Yang[*], William A. Beard[†], Samuel H. Wilson[†], Suse Broyde[‡], Tamar Schlick[*§]

[*] Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute 251 Mercer Street, New York, New York 10012

[†] Laboratory of Structural Biology, National Institute of Environmental Health Sciences, P. O. Box 12233, Research Triangle Park, NC 27709-2233, USA

[‡] Department of Biology, New York University, New York, NY 10003

Running title: Pol β kinetics

Keywords: Dynamics simulations; Human DNA polymerase β; Closed state; Open state; Fidelity

[§]To whom correspondence should be addressed (e-mail:schlick@nyu.edu; fax: 212-995-9152)

March 7, 2001

Abbreviations: pol β, DNA polymerase β; dNTP, 2′-deoxyribonucleoside 5′-triphosphate; ddCTP, 2′, 3′-dideoxyribocytidine 5′-triphosphate; PP_i, pyrophosphate; dRP, deoxyribose phosphate.

I. Supplemental Figures

Fig. S1 Time dependence of conformational parameters for three base-pairs closest to the primer $3'$-terminus. (A)–(D) are sugar pseudorotation angle P and scattergrams of χ/δ for the six base-pair residues, respectively; (E)–(H) are the local base-pair parameters rise, X-displacement, roll, and twist for the three base-pairs. The corresponding values in the canonical fiber models of A-DNA and B-DNA (Leslie et al., 1980) are presented for comparison. The torsion angles χ and δ correspond to $O_{4'}$-$C_{1'}$-N_9-C_4 and $C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (Saenger, 1984). The sugar pseudorotation parameter is defined in (Altona and Sundaralingam, 1972) and the rise, x-displacement, roll, and twist are the nucleic acid structure parameters as defined in (Dickerson et al., 1989). (I) Minor groove (mg) width and depth of primer/template DNA in crystal structures: closed ternary complex and open binary nick complex. (J) Minor groove width/depth of base pair P9-T8 as a function of time from the intermediate structure. P and T indicate primer and template strand, respectively; the subscripts indicate the residue number, $5'$- to $3'$-, in each strand. The minor groove width/depth for canonical X-ray fiber diffraction models of A-DNA and B-DNA (Leslie et al., 1980) is 11.0/2.8 Å and 5.7/7.5 Å, respectively.

Fig. S2 LN performance. (A) Initial model of ternary complex with two bound Mg^{2+}(yellow), 45 Na^+ (green) and 23 Cl^- (purple) ions, as well as 12,418 solvated water molecules (left), error in LN energy components (middle) and LN speedup (right) for DNA pol β complex simulations relative to the explicit Langevin integrator at the single timestep $\Delta t = 1$ fs. LN error is shown for different energy components propagated over 12 ps. (B) Water, DNA, and protein spectra calculated over 12 ps runs and sampled every 2 fs. LN spectra are compared to Langevin spectra as well as spectra from Velocity Verlet. (C) Time evolution of DNA sugar pseudorotation phase angle P and backbone torsion angles of primer residue 11, and time evolutions of ϕ and ψ torsion angles of protein residues Asp192, Glu295, and Tyr296 over 12 ps, respectively. SHAKE was employed for all bonds involving hydrogen atoms in the simulation.

Fig. S3 Pyrophosphate partial charges. (Left) Structure and atom labels; (Right) calculated partial charge for each atom (in esu units).

Fig. S4 Comparison between atomic fluctuations in the MD simulation with X-ray crystallographic B-factors (blue). The simulated atomic fluctuations are calculated over the entire simulation (red), the first (yellow) and second half (green) portion of the trajectory.

II. Supplement to Experimental Procedures

LN Integrator Performance

The LN integrator used here for room-temperature simulations (Schlick et al., 1997; Barth and Schlick, 1998a; Barth and Schlick, 1998b) combines force splitting techniques (Tuckerman et al., 1992) by extrapolation and stochastic coupling for balanced accuracy and stability (Barth and Schlick, 1998a; Schlick and Yang, 2001; Schlick, 2001). Since LN is being used here for a complex biomolecular system, we present here aspects of LN's stability and reliability in terms of thermodynamic, structural and dynamic properties compared to single-timestep Langevin as well as Newtonian (Velocity Verlet) propagators.

The three LN timesteps are related by integers k_1 and k_2 as follows: $\Delta t_m = k_1 \Delta\tau$, $\Delta t = k_2 \Delta t_m$. Thus, the combined force at the reference point X_r, $-\nabla \mathrm{E}(\mathrm{X}_r)$, is evaluated only once per outer timestep Δt; other force terms are updated more frequently, as described above. The splitting of the nonbonded interactions into medium- and long-range parts is accomplished using a smooth switching function. The LN algorithm has been presented in detail elsewhere (Barth and Schlick, 1998a; Barth and Schlick, 1998b; Sandu and Schlick, 1999).

We evaluate LN trajectories through comparison with the traditional small-timestep methods of Langevin (BBK) and Velocity Verlet (VV) in terms of thermodynamic, structural and dynamic properties as a function of time. For this purpose, we fix $\Delta\tau$ to 1 fs, Δt_m to 2 fs, and vary the outer timestep Δt to 6, 20, 40, 80, 100, 120, and 150 fs. All bonds involving hydrogen atoms employed SHAKE both in LN and the reference Langevin method.

Thermodynamic energies calculated over 12 ps of simulation from the equilibrated state are compared with the corresponding energies calculated by the single-timestep Langevin method at $\Delta t = 1$ fs in Fig. S2A, in terms of errors of the energy components and speedup factors relative to the reference Langevin method at $\Delta t = 1$ fs. With varied outer timesteps, bond-length, bond-angle, and torsion energies have the smallest relative errors, around 0.2%. The largest relative error of 2.5% occurs in the van der Waals energy for $\Delta t = 150$ fs, but the relative error of the total energy is still only around 1.2%. For small outer timesteps, relative errors are less than 1%. The higher energy errors for $\Delta t = 20$ fs protocol are likely due to resonance artifacts (Barth and Schlick, 1998b; Sandu and Schlick, 1999). The above results demonstrate the stability of the LN protocol at $\Delta\tau/\Delta t_m/\Delta t$ = 1/2/150 fs and a speedup factor of approximately 4.5 with respect to the reference Langevin at $\Delta t = 1$ fs. Fig. S2

The spectral analysis in Fig. S2B for the protein, DNA, and water atoms obtained from the three algorithms — LN as above, and BBK, at $\gamma = 10$ ps^{-1} and Velocity Verlet ($\gamma = 0$, $\Delta\tau = 1$ fs) — further supports the reliability of this protocol for thermodynamic sampling. Spectral densities are sensitive measures of the algorithmic protocol. The data presented resulted from sampling every 2 fs from the above three 12 ps trajectories. The spectral analysis procedure (Sandu and Schlick, 1999) involves computing the velocity autocorrelation time series for all atoms of interest and obtaining a power spectrum by Fast Fourier Transformation. These spectra are averaged over the water, DNA, and protein atoms separately for the global characterization of the motion.

The LN spectra for water molecules closely resemble the spectra obtained from BBK and VV. For the protein and DNA atoms, although the intensities of both LN and BBK spectra ($\gamma = 10$ ps^{-1}) are weaker than VV spectra intensities ($\gamma = 0$), as expected from Langevin damping, most peak positions calculated from these three algorithms are in good agreement with each other, especially the LN and BBK spectra. Thus, as γ decreases the spectra should become more similar (Barth and Schlick, 1998a; Barth and Schlick, 1998b).

We also analyzed the conformational changes based on the time evolution of ϕ and ψ torsion angles for three protein residues: Asp192, Glu295, and Tyr296, and the DNA backbone torsion angles (χ, α, β, γ, δ) as well as the sugar pseudorotation phase angles, P, for the $3'$-primer terminus guanine residue as shown in Fig. S2C. The time evolutions of these variances using the three algorithms over 12 ps are very similar, respectively. In most cases, the residues exhibit stable oscillations of widths of around 30°. Observed differences in the time evolutions obtained from the different approaches are reasonable and expected.

Finally, we also ascertain the reliability of the simulation comparing the position fluctuations of protein atoms during the production trajectory with values obtained from the experimental temperature factors using the relation $\langle \Delta r_i^2 \rangle = 3B_i/(8\pi^2)$, where Δr_i is the atomic displacement of atom i from its average geometric position and B_i is the corresponding X-ray temperature B–factor (van Gunsteren et al., 1983; Levy et al., 1985; Arnold and Ornstein, 1997). The atomic fluctuations for the backbone atoms of the polymerase are calculated over the entire simulation range as well as the two halves as averages per residue basis. Fig. S4 compares these fluctuations to those obtained from the crystal B–factors of the open complex (Sawaya et al., 1997). Fig. S4

The computed atomic fluctuation patterns show fairly good agreement to the experimental data: the peaks and valleys coincide in general. The regions of higher fluctuations peaks are loop and turn segments positioned between regions of stable secondary structures of α-helices and β-sheets. In particular, the three regions of residues 1–9, 243–249 and 299–305 exhibit relatively large mobility. The intrinsic flexibility (or disorder) of residues 1–9 might explain their absence in the crystallographic structures. Residues 243–249 define the palm's Loop1 as shown in Fig. 2A. Residues 299-305 are part of a loop connecting a β-strand and an α-helix in the thumb. Furthermore, the peak heights for thumb residues 260–335 are larger than those for other subdomains, especially for those calculated over the entire and the first half simulation, which is consistent with that the thumb subdomain has the largest movement in the simulation. Generally, the peak heights for all residues over the second half of the simulation are lower than those over the entire simulation and over the first half. The experimental fluctuations are larger except in three regions (discussed above), which are related to the enzyme's large motion. Differences are also expected given that the simulated protein is free in aqueous solution and the experimental protein is relatively restricted in a crystalline lattice.

References

Altona, C. and Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. J. Amer. Chem. Soc. 94, 8205–8212.

Arnold, G. E. and Ornstein, R. L. (1997). Protein hinge binding as seen in molecular dynamics simulations of native and M61 mutant T4 lysozymes. Biopolymers 41, 533–544.

Barth, E. and Schlick, T. (1998a). Overcoming stability limitation in biomolecular dynamics. I. combining force splitting via extrapolation with Langevin dynamics in LN. J. Chem. Phys. 109, 1617–1632.

Barth, E. and Schlick, T. (1998b). Extrapolation versus impulse in multiple-timestepping schemes. II. linear analysis and applications to Newtonian and Langevin dynamics. J. Chem. Phys. 109, 1633–1642.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kitzing, E. V., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H. and Soumpasis, D. M. (1989). Definitions and nomenclature of nucleic acid structure parameters. EMBO J. 8, 1–4.

Leslie, A. G., Arnott, S., Chandrasekaran, R. and Ratliff, R. L. (1980). Polymorphism of DNA double helices. J. Mol. Biol. 143, 49–72.

Levy, R. M., Sheridan, R. P., Keepers, J. W., Dubey, G. S., Swaminathan, S. and Karplus, M. (1985). Molecular dynamics of myoglobin at 298^o K. results from a 300-ps computer simulation. Biophys. J. 48, 509–518.

Saenger, W. (1984). Principles of Nucleic Acid Structure. Springer-Verlag, New York.

Sandu, A. and Schlick, T. (1999). Masking resonance artifacts in force-splitting methods for biomolecular simulations by extrapolative langevin dynamics. J. Comp. Phys. 151, 74–113.

Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. and Pelletier, H. (1997). Crystal structures of human DNA polymerase β complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry 36, 11205–11215.

Schlick, T. (2001). Time-trimming tricks for dynamic simulations: Splitting force updates to reduce computational work. Structure . (Invited submission of Ways & Means article).

Schlick, T., Barth, E. and Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: bridging the timescale gap between simulation and experimentation. Annu. Rev. Biophys. Biomol. Struct. 26, 181–222.

Schlick, T. and Yang, L. (2001). Long-timestep biomolecular dynamics simulations: ln performance on a polymerase beta/dna system. Multiscale Computational Methods in Chemistry and Physics 00, 00–00.

Tuckerman, M. E., Berne, B. J. and Martyna, G. J. (1992). Reversible multiple time scale molecular dynamics. J. Chem. Phys. 97, 1990–2001.

van Gunsteren, W. F., Berendsen, H. J., Hermans, J., Hol, W. G. and Postma, J. P. (1983). Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data. Proc. Natl. Acad. Sci. U.S.A. 80, 4315–4319.