Supporting Information

Comparative Molecular Dynamics Studies of Human DNA Polymerase η

Melek N. Ucisik and Sharon Hammes-Schiffer*

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801-3364, United States

* To whom correspondence should be addressed: S.H.S.: telephone, (217) 300-0335; email, shs3@illinois.edu.
Figure S1. Systems TTD3'-A (A), TTD3'-G (B), N/A-A (C), and TTD5'-A (D). The protein is shown in cyan, the DNA construct in orange, the incoming nucleotide in ball-and-stick representation, and the TTD lesion or the corresponding normal thymines of the N/A-A system in licorice. The Mg$^{2+}$ ions are displayed as pink spheres. (E) The sequence, secondary structure, and domain information of Pol η, PDB structure 3MR3. The secondary structures were assigned by the STRIDE web server.\textsuperscript{1} The key is as follows: Gray – turn or coil, green – β-sheet, purple – α-helix, blue – 3-10 helix, red – isolated β-bridge. The domain information shown in color-coded sequence was obtained from Dr. Wei Yang of the NIH via private communication. It is in line with their domain descriptions in Ref. 2, where magenta represents the palm, orange represents the finger, cyan represents the thumb, and black represents the little finger domains.
Protocol for Obtaining the Charges for the TTD, dATP, dGTP residues with RESP

The atomic charges for the TTD residue and the free nucleotides dNTP, with “N” either A or G, were obtained using the restrained electrostatic potential (RESP) method.\textsuperscript{3, 4} The transferable nature of the AMBER force fields was utilized in that the atomic charges of the deoxyribose and the phosphate groups were kept as they are in the AMBER libraries for DNA except for the C1’ atoms of the sugar rings.\textsuperscript{5} The H-capped C1’ atoms and the TTD were treated as a new entity, on which RESP calculations were carried out (Figure S2).

![Figure S2](image)

Figure S2. RESP charges were obtained for this molecule consisting of the thymine-thymine dimer and the C1’ atoms that connect the dimer to the deoxyribose sugars. The H atoms shown in red cap this entity in the geometry optimizations and RESP calculations.

Geometry optimizations were performed on the molecule shown in Figure S2 at HF\textsuperscript{6}, B3LYP\textsuperscript{7}, M06L\textsuperscript{8, 9}, and MP2\textsuperscript{10} levels of theory in conjunction with the 6-31G* basis set\textsuperscript{11} in the gas phase using the Gaussian 09 suite of programs. An almost identical optimized structure was obtained from all of these optimizations. Utilizing the HF-optimized geometry, atomic charges were produced to fit the electrostatic potential at points chosen according to the Merz-Singh-Kollman scheme.\textsuperscript{12,13} These charges were optimized with an in-house code over two steps. In the first step, the charges on the more electronegative oxygen and nitrogen atoms were optimized
while keeping the charges on all other atoms identical to the AMBER charges. In the second step, the charges of the oxygens and nitrogens were fixed to their optimized values from the previous step, whereas the charges on all other atoms were optimized. In this step, equivalencies were defined for the atoms that are not in the immediate vicinity of the cyclic region of the molecule. The charges on the three hydrogens of the methyl groups were also set to be identical. Note that the charges on the C1’ atoms were fixed to their AMBER values throughout the procedure. The charges on the capping hydrogens were adjusted to ensure that the total charge on the rest of this neutral molecule would be \[2 + (\text{the total charge on the two sugar rings and the phosphate groups as they are found in the AMBER libraries})\]. The bond, angle, and dihedral parameters were obtained using the parmchk2 utility of the AmberTools13 employing the ff10 force field parameters.

The quality of the RESP charges determined for the TTD base piece (Figure S2) was tested through a comparison of the optimized geometries and dipole moments at the HF/6-31G* and MM levels of theory. The optimized geometries yielded good structural agreement with a root-mean-squared-deviation (RMSD) of 0.29 Å. The dipole moment obtained at the HF/6-31G* level of theory has a total magnitude of 7.22 D with \(x\)-, \(y\)-, and \(z\)-components of -2.40, -2.40, and -6.37 D, whereas the dipole moment computed at the MM level for the same Cartesian coordinates (i.e., for the geometry optimized at the QM level) has a total magnitude of 7.21 D with the \(x\)-, \(y\)-, and \(z\)-components of -2.29, -2.46, and -6.38 D. The scalar product of the MM and QM dipole moments was 1.00 after normalization, indicating that these dipole moments are oriented in the same direction.

The dATP and dGTP molecules were treated similarly to obtain their RESP charges. This approach was also utilized by Meagher et al. in previous work for calculating the RESP charges.
of the ATP molecule. The charges on the triphosphate were re-optimized, while the AMBER charges were employed for the adenine base and the deoxyribose sugar. The triphosphate was capped with a methyl group to yield the molecule shown in Figure S3 with a molecular charge of $-4$.

![Figure S3](image)

**Figure S3.** RESP charges calculated for the atoms shown in black in the dNTP (N=A or G) molecule. The methyl group shown in red serves as the capping group.

The optimized structure was obtained from Meagher et al., and the atomic charges were computed to fit the electrostatic potential at points selected according to the Merz-Singh-Kollman scheme. The total charge on the capping group was set equal to the total charge of the 3’-end deoxyribose and the base, either A or G, whereas the charge on the C5’ atom was chosen to be identical to the charge of C5’ in a 3’-A or 3’-G nucleotide in the AMBER libraries for DNA (nucleic12.lib). For phosphorus and oxygen atoms, the atom types introduced by Meagher et al. were adapted, and equivalencies for the terminal oxygens on the first, second, and third phosphorus atoms were defined. Unlike the TTD case, all of these RESP charge optimizations were performed at the same time. The bond, angle, and dihedral parameters gathered with the aid of the parmchk2 tool of AmberTools using the ff10 AMBER parameters and were combined with the parameters acquired by Meagher et al. in the previous work on ATP.
Additional Remarks about Structure Preparation

The topologies for the TTD3’-A and TTD3’-G simulations were based on the PDB structure 3MR3, the N/A-A topology was based on the PDB structure 3MR2, and the TTD5’-A topology was based on the PDB structure 3SI8. The missing residues in each of these crystal structures were filled in with Schrödinger’s Maestro program. The glycerol molecules, the second free nucleobase, and all waters except the one bound to one of the Mg$^{2+}$ ions (number 439 for 3MR3, 463 for 3MR2, and 495 for 3SI8) were removed from the PDB structure prior to the simulations.

Figure S4. RMSDs of the C$_\alpha$ atoms for the three independent simulations associated with the systems TTD3’-A (A), TTD3’-G (B), N/A-A (C), and TTD5’-A (D). Each color represents data obtained from one of the three independent trajectories.
Figure S5. RMSFs of the three independent simulations associated with the systems TTD3’-A (A), TTD3’-G (B), N/A-A (C), and TTD5’-A (D). Each color represents data obtained from one of the three independent trajectories.
Figure S6. Surface areas for the residues involved in the nucleotidyl addition reaction, D13, M14, D115, and E116 (denoted with “rxn center”), and the residues interacting with the DNA around the active site, F18, Q38, Y39, I48, R61, S62, K86, L89, Y92, and R93 (termed “active site”). These data have been extracted from the first MD trajectories for each system.

Figure S7. Surface areas for the reaction centers (lower curves) and the residues that interact with the reaction center (upper curves) for all three independent trajectories of the four systems: TTD3’-A (A), TTD3’-G (B), N/A-A (C), and TTD5’-A (D). Each color represents data obtained from one of the three independent trajectories.
Figure S8. Time evolution of the radius of gyration (lower curves) and the largest distance between any two protein atoms (upper curves), for the three independent trajectories for systems TTD3’-A (black), TTD5’-G (red), N/A-A (green), and TTD5’-A (blue). Each color represents data obtained from one of the three independent trajectories.
Figure S9. Cross-correlation maps for the systems TTD3'-A (a, b, c), TTD3'-G (d, e, f), N/A-A (g, h, i), and TTD5'-A (j, k, l). Color scheme: $-1$/red represents residues that move in an anticorrelated fashion (i.e., opposite directions) and $1$/blue represents residues that move in a fully correlated fashion (i.e., same direction).
Figure S10. The systems TTD3’-A, TTD3’-G, N/A-A, and TTD5’-A overlaid to compare the locations of their DNA sequences. Different colors represent different nucleotides. The three of the same color that are on top of each other correspond to the DNA constructs in systems TTD3’-A, TTD3’-G, and N/A-A. In the fourth system, TTD5’-A, the DNA piece is shifted upstream by one base pair.
Figure S11. Time evolution of the major (black) and minor (red) groove widths for the three independent trajectories of the systems TTD3'-A (row A), TTD3'-G (row B), N/A-A (row C), and TTD5'-A (row D).

Figure S12. Histograms depicting the number of hydrogen bonds formed within the nucleic acids for the three independent trajectories of the systems TTD3'-A (A), TTD3'-G (B), N/A-A (C), and TTD5'-A (D). The dashed and solid lines represent the data collected with a 3.0 Å and 3.5 Å cutoff, respectively, for the definition of a hydrogen bond. The donor-hydrogen-acceptor angle cutoff was 135° for all cases. Each color represents data obtained from one of the three independent trajectories.
Figure S13. Quantification of the most common amino acid hydrogen-bonding partners of the dNTP molecules in the systems TTD3’-A (black), TTD3’-G (red), N/A-A (green), and TTD5’-A (blue).
Figure S14. RMSDs of the dNTP for the three independent trajectories of the systems TTD3’-A (A), TTD3’-G (B), N/A-A (C), and TTD5’-A (D). Each color represents data obtained from one of the three independent trajectories.
Figure S15. RMSDs of the TTD and TT motifs for the three independent trajectories of the systems TTD3'-A (A), TTD3'-G (B), N/A-A (C), and TTD5'-A (D). Each color represents data obtained from one of the three independent trajectories.
Figure S16. Histograms for the nucleophilic attack distance between $P_\alpha$ of the dNTP molecule and the O3' atom of the 3'-end of the DNA primer for the three independent trajectories (designated with colors black, red, and green) of the systems TTD3'-A (A), TTD3'-G (B), N/A-A (C), and TTD5'-A (D).
Figure S17. RMSFs of the P atoms in the DNA constructs for the three independent trajectories of the systems TTD3′-A (A), TTD3′-G (B), N/A-A (C), and TTD5′-A (D). Each color represents data obtained from one of the three independent trajectories. The newly forming primer strands are represented as straight lines, and the template strands are represented as dashed lines. In (A), (B), and (D), the P atoms within the TTD dimer are labeled with P1 and P2.
Table S1. The nucleic acid residues in the four systems and their residue numbers.

| Template (T) or Primer (P) | # Nucleotides | Nucleotide Residues and their Numbers |
|---------------------------|---------------|----------------------------------------|
|                           |               | 445 446 447 448 449 450 451 452 453 454 455 456 457 |
| 3MR2                      |               |                                        |
| T                         | 13            | DT DC DA DT DT DA DT DG DA DC DG DC DT |
| P                         | 9             | DT DA DG DC DG DT DC DA DT             |
|                           |               | 436 437 438 439 440 441 442 443 444    |
| 3MR3                      |               |                                        |
| T                         | 11            | DC DA TTD DA DT DG DA DC DG DC DT      |
| P                         | 9             | DC DA DG DC DG DT DC DA DT             |
|                           |               | 436 437 438 439 440 441 442 443 444    |
| 3S18                      |               |                                        |
| T                         | 12            | DT DA DA DC TTD DA DT DG DA DC DG DC   |
| P                         | 9             | DT DG DC DG DT DC DA DT DA             |
|                           |               | 436 437 438 439 440 441 442 443 444    |
**Table S2.** Residues that are within 5 Å of the dNTP molecule or the TTD lesion in selected configurations from the TTD3’-A, TTD3’-G, and TTD5’-A simulations. A consensus active site was defined by including all of these residues and was utilized in the surface area calculations.

| System       | Residues that are less than 5 Å away from the dNTP or the TTD |
|--------------|---------------------------------------------------------------|
| TTD3’-A      | D13, M14, D15, C16, F17, F18, V19, Q38, Y39, I48, A49, V50, Y52, R55, W64, A87, I114, D115, E116, K231, K323, N324, P326, L378 |
| TTD3’-G      | D13, M14, D15, C16, F17, F18, V19, Q38, Y39, I48, A49, V50, R55, S62, W64, A87, L89, S113, I114, D115, E116, K231, K323, N324, F325, P326, R371, L378, M421 |
| TTD5’-A      | D13, M14, D15, C16, F17, F18, V19, Q38, K40, G46, I48, A49, V50, Y52, R55, R61, A87, L89, R93, S113, I114, D115, E116, K231, S322, K323, N324, R371, L378, M421, F423 |
| Consensus active site | D13, M14, D15, C16, F17, F18, V19, Q38, Y39, K40, G46, I48, A49, V50, Y52, R55, R61, S62, W64, A87, L89, R93, S113, I114, D115, E116, K231, S322, K323, N324, F325, P326, R371, L378, M421 |
Table S3. RMSF values determined for the dNTP molecule and the TTD/TT motif from all MD trajectories. The italicized RMSF values for the N/A-A system correspond to the thymine that does not form base-pairing interactions with the dNTP molecule. The average RMSF for each system is given in bold. The greatest mobility for dNTP is observed in the TTD3’-G system, which includes a dGTP instead of a dATP, and the greatest mobility for the TTD/TT motif is observed in the N/A-A system, which does not include a TTD.

| System   | Trajectory | RMSF (Å) of dNTP | RMSF (Å) of TTD/TT |
|----------|------------|------------------|--------------------|
| TTD3’-A  | 1          | 0.226            | 0.568              |
|          | 2          | 0.236            | 0.481              |
|          | 3          | 0.226            | 0.453              |
| Average  |            | **0.229**        | **0.501**          |
| TTD3’-G  | 1          | 0.322            | 0.678              |
|          | 2          | 0.236            | 0.539              |
|          | 3          | 0.360            | 0.538              |
| Average  |            | **0.306**        | **0.585**          |
| N/A-A    | 1          | 0.341            | 1.451              |
|          |            |                  | 0.975              |
|          | 2          | 0.256            | 1.340              |
|          |            |                  | 0.831              |
|          | 3          | 0.221            | 0.734              |
|          |            |                  | 0.472              |
| Average  |            | **0.273**        | **0.967**          |
| TTD5’-A  | 1          | 0.254            | 0.321              |
|          | 2          | 0.254            | 0.346              |
|          | 3          | 0.260            | 0.475              |
| Average  |            | **0.256**        | **0.381**          |
References:

1. Heinig, M.; Frishman, D., Stride: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res.* 2004, 32, W500-W502.

2. Biertumpfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and Mechanism of Human DNA Polymerase η. *Nature* 2010, 465, 1044-1048.

3. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A., Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins. *J. Comput. Chem.* 1995, 16, 1357-1377.

4. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of Resp Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *J. Am. Chem. Soc.* 1993, 115, 9620-9631.

5. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force-Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 1995, 117, 5179-5197.

6. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*. Second ed.; John Wiley & Sons Ltd.: The Atrium, Southern Gate, Chichester, West Sussex, England, 2004.

7. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J. Phys. Chem.* 1994, 98, 11623-11627.

8. Zhao, Y.; Truhlar, D. G., A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* 2006, 125, 194101.

9. Zhao, Y.; Truhlar, D. G., The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* 2008, 120, 215-241.
10. Head-Gordon, M.; Pople, J. A.; Frisch, M. J., MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503-506.
11. Ditchie, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. 9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.
12. Singh, U. C.; Kollman, P. A., An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129-145.
13. Besler, B. H.; Merz, K. M.; Kollman, P. A., Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431-439.
14. Meagher, K. L.; Redman, L. T.; Carlson, H. A., Development of Polyphosphate Parameters for Use with the Amber Force Field. *J. Comput. Chem.* **2003**, *24*, 1016-1025.
15. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *Amber 14*. University of California, San Francisco: 2014.
16. Maestro Version 9.3; Schrödinger, LLC: New York, NY, 2012.