

# Influence of local sequence context on damaged base conformation in human DNA polymerase $\iota$ : molecular dynamics studies of nucleotide incorporation opposite a benzo[a]pyrene-derived adenine lesion

Kerry Donny-Clark and Suse Broyde\*

Department of Biology, New York University, New York, NY 10003, USA

Received July 6, 2009; Revised and Accepted August 24, 2009

## ABSTRACT

Human DNA polymerase  $\iota$  is a lesion bypass polymerase of the Y family, capable of incorporating nucleotides opposite a variety of lesions in both near error-free and error-prone bypass. With undamaged templating purines polymerase  $\iota$  normally favors Hoogsteen base pairing. Polymerase  $\iota$  can incorporate nucleotides opposite a benzo[a]pyrene-derived adenine lesion (dA\*); while mainly error-free, the identity of misincorporated bases is influenced by local sequence context. We performed molecular modeling and molecular dynamics simulations to elucidate the structural basis for lesion bypass. Our results suggest that hydrogen bonds between the benzo[a]pyrenyl moiety and nearby bases limit the movement of the templating base to maintain the *anti* glycosidic bond conformation in the binary complex in a 5'-CAGA\*TT-3' sequence. This facilitates correct incorporation of dT via a Watson–Crick pair. In a 5'-TTTA\*GA-3' sequence the lesion does not form these hydrogen bonds, permitting dA\* to rotate around the glycosidic bond to *syn* and incorporate dT via a Hoogsteen pair. With *syn* dA\*, there is also an opportunity for increased misincorporation of dGTP. These results expand our understanding of the versatility and flexibility of polymerase  $\iota$  and its lesion bypass functions in humans.

## INTRODUCTION

Various mutagenic DNA adducts can compromise replication fidelity, causing mutations in proto-oncogenes or tumor suppressors that can lead to cancer (1,2).

Often these adducts block the replicative polymerase, and low-fidelity bypass polymerases are called upon to extend past the distorting DNA damage (3–6). One such error-prone bypass polymerase is human DNA polymerase  $\iota$  (pol $\iota$ ), a member of the Y-family of damage bypass polymerases (5,7–9). Pol $\iota$  is unique among DNA polymerases discovered to date in its exceptionally low fidelity and processivity (8,10,11). In addition, pol $\iota$  is the only polymerase that shows greatly different error rates dependent upon the templating base when replicating undamaged DNA. dT is inserted opposite templating dA with relatively high accuracy, while dG is preferred to dA opposite templating dT (8,10,12). Though pol $\iota$  is unique in many ways, its biological role remains a subject of ongoing investigation (13).

The first crystal structure of pol $\iota$  showed that Hoogsteen base pairing is used in the active site of the pol $\iota$  ternary complex, comprised of the enzyme, an undamaged dA as the templating base, and an incoming dTTP (14). Ternary structures with templating dG and binary complexes with templating dA and dG without incoming dNTP soon followed, showing Hoogsteen base pairing (templating base *syn*) in the ternary structures and *anti* templating bases in the binary structures (15,16). These structures suggested a mechanism in which steric hindrance caused by the incoming nucleotide forces templating purine bases to rotate from the *anti* to the *syn* glycosidic bond conformation in order to form a Hoogsteen base pair (15). Recently, the first pol $\iota$  crystal structures with a templating *anti* dT showed that the incoming dATP can then adopt the *syn* conformation instead (17). The polymerase active site holds the sugar of the templating base in a hydrophobic pocket formed by residues Gln59, Lys60 and Leu62. This limits the space between the sugars of the templating base and the incoming dNTP such that a Watson–Crick base pair, with a C1' to C1' distance of  $\sim 10.6$  Å (18), cannot be

\*To whom correspondence should be addressed. Tel: +1 212 998 8231; Fax: +1 212 995 4015; Email: broyde@nyu.edu

accommodated without an adjustment of the polymerase. A Hoogsteen pair has a shorter C1' to C1' distance of ~8.6 Å (18), and can fit in the active site of the ternary complex without steric clashes.

Hoogsteen base pairing with templating dG reverses the position of the N<sup>2</sup> and C8 atoms, both of which are sites for covalent linkage of DNA adducts; N<sup>2</sup> moves from the minor groove side to the major groove side of the nascent B-DNA helix, while C8 moves from the major groove side to the minor groove side. Polt ternary complexes show that the major groove side of the nascent helix is spacious and solvent-exposed, while there is little space between the polymerase and the DNA on the minor groove side (14–17,19). Thus, Hoogsteen base pairing would move guanine N<sup>2</sup> adducts from the cramped minor groove side of the helix to the spacious major groove side. In primer extension experiments with a variety of dG-N<sup>2</sup> adducts, polt is able to incorporate the correct partner opposite the damaged base with varying efficiencies (20–24), supporting this hypothesis. A crystal structure of a ternary complex of polt with a dG-N<sup>2</sup> adduct shows the N<sup>2</sup>-ethylguanine adduct placed on the spacious major groove side (24), and biochemical experiments provide further support for this hypothesis (25,26). Polt may deal with bulky adducts on C8 of dG in an analogous fashion, accommodating a conformation that keeps bulky lesions away from the cramped minor groove side of the protein. Simulations from our laboratory suggest that a dG-C8 acetylaminofluorene adduct is well positioned on the spacious major groove side with Watson–Crick base pairing between the damaged template and the incoming dCTP (27).

Adenine adducts present a different case. In the case of dA-N<sup>6</sup> adducts, the adduct is placed on the spacious major groove side in both Hoogsteen and Watson–Crick base pairs. Ethenoadenine adducts that interfere with Watson–Crick bonding have been shown to form a Hoogsteen pair with incoming dTTP in a polt ternary complex (28). However, N3-methyladenine lesions in polt may be unable to form a Hoogsteen pair due to a clash between the methyl group and the DNA backbone (29); yet polt can bypass this lesion with incorporation of dT (30), leaving open the possibility that a Watson–Crick base pair may be employed. Recent kinetic analyses utilizing incoming dNTPs containing base analogs suggest that templating dC may use Watson–Crick base pairing to select dGTP (31). Thus polt may be able to use either Hoogsteen or Watson–Crick base pairing with differing templates, lesions and incoming dNTPs.

Benzo[*a*]pyrene (BP) is a well-studied model pre-carcinogen found in by-products of incomplete combustion, including tobacco smoke and automobile exhaust among others (1). Mirror image pairs of (±)*syn*- and (±)*anti*-benzo[*a*]pyrene diol epoxides (BPDEs) are products of BP metabolic activation through the predominant diol epoxide metabolic activation pathway (32), though other pathways have been identified (33,34). These diol epoxides can then react by *cis*- or *trans*-epoxide ring opening, covalently bonding to the exocyclic N<sup>2</sup> group of guanine or N<sup>6</sup> group of adenine (35–37). Here, we have investigated adenine lesions derived from

*trans* opening of 7*r*,8*t*-dihydroxy-*t*9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene [(+)*anti*-BPDE], the most frequent product of the BPDE metabolic activation pathway as well as the most carcinogenic metabolite (32,38). This results in the 10*S*(+)-*trans-anti*-[BP]-N<sup>6</sup>-dA (BP-dA) adduct, the subject of this study (Figure 1A). This adduct has been demonstrated to be mutagenic in *E. coli* (39,40), and to be inducing miscoding in primer extension studies with various DNA polymerases (20,40,41).

The present study was motivated by an interest in investigating the relatively error-free bypass of BP-dA by polt observed in certain sequence contexts. Frank *et al.* (41) have investigated the 10*S*(+)-*trans-anti*-[BP]-N<sup>6</sup>-dA lesion (among others) in primer extension assays with polt in two sequence contexts: SeqI (template 5'-CAGA\*TTTA GAGTCTGC-3'), a mutational hotspot from the *E. coli supF* gene, and SeqII (template 5'-TTTA\*GAGTCTGCT CCC-3'), a mutational cold spot from the same gene (Figure 1B) (41,42). A\* represents the modified adenine. While polt predominantly incorporates the correct partner dT in both SeqI and SeqII, SeqII misincorporates dG with ~14-fold higher efficiency than SeqI (41). In both sequences, extension beyond the lesion is severely inhibited, though addition of polk substantially increases extension (41). This inability to extend from a lesion after incorporating a base opposite it is a hallmark of polt, with other studies also showing the requirement for a second polymerase to extend from various lesions (12,21,43). In a 5'-GACA\*AG-3' sequence context, polt appears to be more mutagenic during bypass of the BP-dA adduct than in the 5'-TTTA\*GA-3' and the 5'-CAGA\*TT-3' sequences; however, the kinetic parameters have not been evaluated for this case (20). In the present study, we investigate the structural underpinnings of the relatively error-free bypass of a BP-dA lesion by polt in the two sequences employed in Frank *et al.* (41) using molecular modeling and 20 ns of molecular dynamics (MD) simulations.

We constructed initial models of the damaged BP-dA ternary complexes in each sequence with two sets of α', β' orientations for the lesion (Figure 1A), both *anti* and *syn* glycosidic bond templating dA conformations, and all four incoming dNTPs (Supplementary Table S1). We also simulated unmodified *syn* dA in both sequence contexts, with all 4 incoming dNTPs, as well as unmodified *anti* dA with incoming dTTP. Finally, in order to investigate the state of the active site prior to entrance of the dNTP, we simulated binary complexes with *anti* templates as observed in the binary complex crystal structures (15), with and without the BP lesion, in both sequences. The binary complex BP initial structures were modeled with both α', β' orientations of the lesion employed in the damaged ternary complex simulations above. Thus, we ran 56 simulations in total.

In the SeqI undamaged control binary complex simulation with *anti* template, we observed several hydrogen bonds that maintain a normal binary complex active site with the templating base anchored in the *anti* conformation. This facilitates correct incorporation of dT via a Watson–Crick pair. The SeqII undamaged control

binary complex simulation with *anti* template shows the damaged templating base shifted toward the major groove, which would lower the barrier to rotation around the glycosidic bond to a *syn* conformation, facilitating incorporation of dT via a Hoogsteen base pair. Furthermore, this more-open binary complex active site might explain the observed enhanced misincorporation of dG relative to the other sequence. These results further expand our understanding of the versatility and flexibility of polymerase  $\epsilon$  and its possible lesion bypass functions in humans.

## MATERIALS AND METHODS

### Molecular modeling of initial unmodified control structures for MD

**Ternary complexes.** We used the polt ternary complex crystal structure containing a dA-dTTP Hoogsteen base pair [PDB (44) ID: 2FLL (15)] as the basis for our initial *syn* dA unmodified models. Missing loops containing residues 371 to 378 and residues 395 to 403 were modeled with the program 'Modeller' (45,46) on the Modloop web server (47). The DNA sequences were taken from primer extension experiments performed in Frank *et al.* (41) and modeled into the crystal structure using the InsightII 2005 software package (Accelrys Inc.). The coordinates with PDB ID: 2FLL contained 8 residues on the templating strand and 6 residues on the primer strand. To model the Frank *et al.* sequences (41), 16 template residues and 12 primer residues were required (Figure 1B). The added residues were modeled into the crystal structure with standard B-DNA geometry. The base of the incoming dNTP was remodeled in order to create starting models for all four incoming dNTPs in each sequence. These structures were used as our unmodified controls for the ternary complex simulations. *Anti* dA unmodified initial structures were constructed by replacing the BP lesion from the most representative structure of the last cluster of the modified *anti* BP-dA/dTTP simulations with a hydrogen (See below for details of clustering, most representative structures and BP-dA simulations).

**Binary complexes.** Binary unmodified control structures were constructed by removing the incoming nucleotide from the dA-dTTP unmodified control initial structures for each sequence and changing the glycosidic torsion of the templating dA to *anti*.

### Molecular modeling of initial BP-dA structures for MD

**Ternary complexes.** The initial unmodified control structures with *syn* template (details above) for both sequences and all four incoming dNTPs were used as the basis for initial BP-dA structures. The glycosidic bond conformation of the BP-dA was *syn*, and all incoming dNTPs retained the *anti* glycosidic bond conformation of the crystal structure. Next, a BP moiety was modeled into the unmodified dA-dTTP structures for each sequence and a conformational search was performed to find optimal  $\alpha'$  and  $\beta'$  linkage torsion angles (Figure 1B)

for the initial structures with minimal collisions. For this purpose, the  $\alpha'$  and  $\beta'$  torsion angles were surveyed over their 360° range at 10° intervals starting at 5°, for a total of 1296 structures created for each sequence. These structures were evaluated for steric clashes with the bumpcheck utility of InsightII (Accelrys Inc.), and the structures with the fewest close contacts in each sequence were selected. In order to fully explore the possible conformations of the ternary complexes, we created a second set of structures with the glycosidic torsion angle  $\chi$  of the templating dA in the *anti* conformation (Supplementary Table S1) and repeated the search for optimal structures. In this case, both *syn* and *anti* incoming dNTP conformations were investigated. Specifically, pyrimidine dNTPs were only studied in the *anti* conformation, since pyrimidines rarely adopt the *syn* domain (48), while both *syn* and *anti* purine dNTPs were examined. While the optimal  $\alpha'$ ,  $\beta'$  combinations differed in the 5'-GA\*TT-3' sequence (SeqI) and the 5'-TTTA\*GA-3' sequence (SeqII), for more thorough exploration of the conformational possibilities for the BP moiety, initial models were constructed for both sets of  $\alpha'$ ,  $\beta'$  combinations in each sequence; this was done for both *syn* and *anti* damaged templates (Supplementary Table S1). Thus, a total of (2 sequences  $\times$  4 incoming *anti* dNTPs  $\times$  2 initial  $\chi$  values  $\times$  2 initial BP  $\alpha'$ ,  $\beta'$  combinations + 2 sequences  $\times$  2 incoming *syn* purines  $\times$  2 initial BP  $\alpha'$ ,  $\beta'$  combinations) 40 initial modified ternary complexes were constructed for the MD simulations.

**Binary complexes.** The binary complex BP-dA modified initial structures were created by adding BP moieties to the unmodified binary controls with the 2 initial sets of BP  $\alpha'$ ,  $\beta'$  combinations used in the ternary complexes, for a total of (2 sequences  $\times$  2 initial BP  $\alpha'$ ,  $\beta'$  combinations) 4 initial modified BP-dA binary complexes (Supplementary Table S1).

Including both *syn* template and *anti* template unmodified controls and BP-dA structures 50 ternary initial structures and 6 binary initial structures were modeled and 20 ns MD performed.

### Ternary complex BP-dA trajectory analysis

Because of the large number of ternary complex BP-dA simulations, we utilized a three-stage procedure to analyze the structures: (i) we analyzed all 40 BP-dA ternary complex trajectories for  $Mg^{2+}$  coordination, base stacking energy estimates,  $\chi$ ,  $\alpha'$  and  $\beta'$  torsion angle values, hydrogen bond occupancies between the BP-dA and the enzyme, DNA, and incoming dNTP, C1'-C1' distance, P $\alpha$ -O3' distance, and P $\alpha$ -O3' attack angle; (ii) the binary complex results showed that SeqI would maintain an *anti* orientation for the templating base due to several hydrogen bonds involving the BP-dA, one of which is sequence specific. In SeqII the templating BP-dA is more flexible, and thus able to sample both *syn* and *anti* conformations. Therefore, we further analyzed only *anti* template simulations for SeqI, and both *anti* and *syn* simulations for SeqII for each dNTP/sequence combination; (iii) we then selected one most favored trajectory

from the *anti* SeqI and the *syn* and *anti* SeqII simulations for each incoming dNTP/sequence combination; this selection was based upon the number and occupancy of hydrogen bonds between the templating BP-dA and the incoming dNTP and using the C1'-C1' distance as a tiebreaker, with preference given to shorter distances. This selection was guided by the fact that Y-family polymerases such as polt have a relatively spacious active site and therefore use hydrogen bonding to promote fidelity (6,49,50), and also that polt encourages a short C1'-C1' distance (15). Selections are shown in Supplementary Table S2. Note that all purine-purine mismatches involve a *syn/anti* pairing, as is to be expected from polt's preference for a short C1'-C1' distance. The structures selected by this three-stage procedure were then clustered (see below). The most representative structures (see below) for each cluster within the trajectories were obtained and further analyzed visually. Full details of all stages of these analyses are presented below. In addition, active site figures and brief descriptions of BP-dA ternary complex structures with *syn* templates in the 5'-CAGA\*TT-3' sequence (SeqI) and *anti* templates in the 5'-TTTA\*GA-3' sequence (SeqII) are presented as Supplementary Figures S31-S38. For these figures, we selected the most favored *syn* structure for each incoming dNTP in the 5'-GA\*TT-3' sequence (SeqI), and the most favored *anti* structure for each incoming dNTP in the 5'-TA\*GA-3' sequence (SeqII). These figures are of the active sites from the most representative structure of the last cluster of each simulation.

### Force field

The Cornell *et al.* force field (51) with modifications (52,53) and the parm99sb parameter set (54,55) modified by parmbsc0 (56) were employed for all simulations. For the BP-dA modified nucleoside partial charge calculations, a multiconformational protocol was employed using all of the initial models'  $\alpha'$ ,  $\beta'$  and  $\chi$  combinations as inputs (Figure 1B, Supplementary Table S1) (57,58). Partial charges for the BP-dA were calculated with the HF method and the 6-31G\* basis set (59) using Gaussian 03 (60), and the restrained electrostatic potential fitting algorithm RESP (61,62) was employed to fit the charge to each atom center (Supplementary Table S3). All dNTP charges were taken from previous work (63). Parameters for atom types not found in the PARM99 parameter set were taken from the GAFF (64) parameter set or developed by analogy to chemically similar atom types in the PARM99 and GAFF parameter sets (Supplementary Table S3).

### MD protocol

All minimizations and MD simulations used the PMEMD module of the 'Amber 9' software suite (65). The LEaP module of 'Amber 9' was used to add hydrogen atoms and neutralize the system with sufficient Na<sup>+</sup> atoms to bring the net charge to zero. Hydrogen atoms of the solute (DNA, polymerase and incoming dNTP) were minimized with implicit solvent using a distance-dependent dielectric

function of  $\epsilon = 4.0r$  (where  $r$  is the distance between an atom pair) for 600 steps of steepest descent, followed by 600 steps of conjugate gradient. The resulting ternary complex was reoriented with SIMULAID (66) to minimize the number of water molecules needed to solvate the system. A periodic TIP3P (67) rectangular water box with a buffer distance of 10 Å between each wall and the closest solute atom in each direction was added with the LEaP module of 'Amber 9'. Box dimensions were approximately 80 Å × 80 Å × 95 Å, with a total of ~14400 water molecules.

All systems employed the following equilibration and MD protocols: (i) minimization of the counterions and solvent molecules for 2500 steps of steepest descent and 2500 steps of conjugate gradient, with 50 kcal/mol restraints on the solute atoms; (ii) 30 picoseconds (ps) initial MD at 10 K with 25.0 kcal/mol restraints on the solute to allow the solvent to relax; (iii) heat-up from 10 to 310 K (37°C) at constant volume over 30 ps with 10 kcal/mol restraint on the solute; (iv) 30, 40 and 50 ps MD at 1 atm and 310 K with decreasing restraints of 10, 1 and 0.1 kcal/mol, respectively, on solute atoms; and (v) production MD was conducted at 1 atm, 310 K for 20 ns, with 1 ps coupling constants for both pressure and temperature.

In all MD simulations, long-range electrostatic interactions were treated with the particle mesh Ewald method (68,69). A 9 Å cutoff was applied to the non-bonded Lennard-Jones interactions. The SHAKE algorithm (70) was applied to constrain all bonds involving hydrogen atoms with a relative geometrical tolerance of 10<sup>-5</sup> Å. Langevin dynamics with the collision frequency  $\gamma$  set to 5.0 ps<sup>-1</sup> (71) were used for equilibration (steps ii-iv, above), and the Berendsen coupling algorithm (72) was used for temperature scaling in the production MD simulations (step v, above). A 2-femtosecond (fs) time step was used, and the translational/rotational center-of-mass motion was removed every 0.5 ps (73).

All simulations showed reasonable stability after ~2.5 ns. This was evaluated by inspecting the RMSD of the active site (defined as all residues having atoms with 8.0 Å of any atom in the nascent base pair) as compared to the initial state after equilibration over time (Supplementary Figures S1D-S22D). Therefore, all analyses presented here were performed on the final 17.5 ns of the trajectories. The ptraj module of 'Amber 9' (65) was used to collect ensemble average values of distances and angles, and the anal module of 'Amber 6' (74) was used to evaluate the van der Waals interaction energies as an estimate of base stacking energies; these were evaluated for the incoming dNTP with the primer terminal base (base 17) and the templating base with base 12 (Figure 1B).

### Trajectory clustering

The algorithm we implemented to cluster our trajectories has been previously described as quality threshold clustering in the context of clustering DNA sequences (75), and has also been employed in *de novo* protein structure prediction (76). This algorithm places all structures that are

within a certain RMSD cutoff of each other into a cluster. The cluster with the most members is kept, and all members of that cluster are taken out of consideration in further rounds to ensure that each frame is in only one cluster. This process is repeated until the largest cluster has less than 500 members.

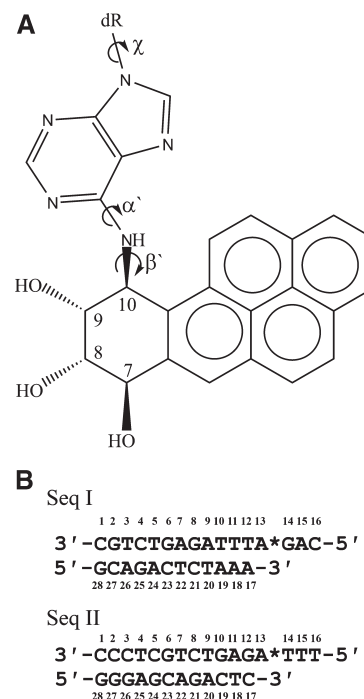
We have selected this algorithm for the purpose of clustering frames from MD trajectories, a new application as far as we are aware. The algorithm does not require the user to supply the number of clusters *a priori*. In addition, not all structures are placed in a cluster. If a given structure is too far from all other structures, it will remain outside of all other clusters. This should allow us to form clusters that represent the troughs of the energy landscape. This combination of features is particularly useful for analyzing MD simulations, providing a valuable addition to other available algorithms used for this purpose (77,78).

### Most representative structures

The most representative structures for each cluster are the structures that have the lowest active site RMSD to all other structures in the cluster. These most representative structures for each cluster were collected for each trajectory and compiled into the animations available as Supplementary Movies S1–S22. For illustration purposes only, the most representative structures from the last cluster in each trajectory (or one nearby where specified) were used for Figures 2–4 and Supplementary Figures S1–S38. However, full-scale trajectory analyses were performed for the entire ensembles from 2.5 ns to 20 ns for those simulations that met our selection criteria.

## RESULTS

We performed simulations with both damaged and undamaged binary and ternary complexes and both *syn* and *anti* templating dA/BP-dA. Damaged BP-dA ternary complex simulations included all four incoming dNTPs, and both *syn* and *anti* glycosidic bond conformations for the templating base as well as two different initial  $\alpha'$ ,  $\beta'$  combinations (Supplementary Table S1). Pyrimidine dNTPs were only studied in the *anti* conformation, since pyrimidines rarely adopt the *syn* domain (48), while both *syn* and *anti* purine dNTPs were examined. All simulations were performed in both sequence contexts (Figure 1B) with all other initial conditions identical. Damaged BP-dA binary complex simulations with *anti* templating BP-dA were performed with both initial  $\alpha'$ ,  $\beta'$  combinations used in the ternary complex BP-dA simulations in both 5'-GA\*TT-3' (SeqI) and 5'-TA\*GA-3' (SeqII) sequence contexts. We also performed unmodified binary control simulations with *anti* templating dA, and unmodified ternary control simulations with *syn* templating dA with all four incoming dNTPs as well as *anti* templating dA with incoming dTTP; these simulations were performed in both sequence contexts. A total of 56 simulations were performed. Simulations were equilibrated and then run for 20 ns of production MD, and analyzed as described in 'Materials and Methods' section.



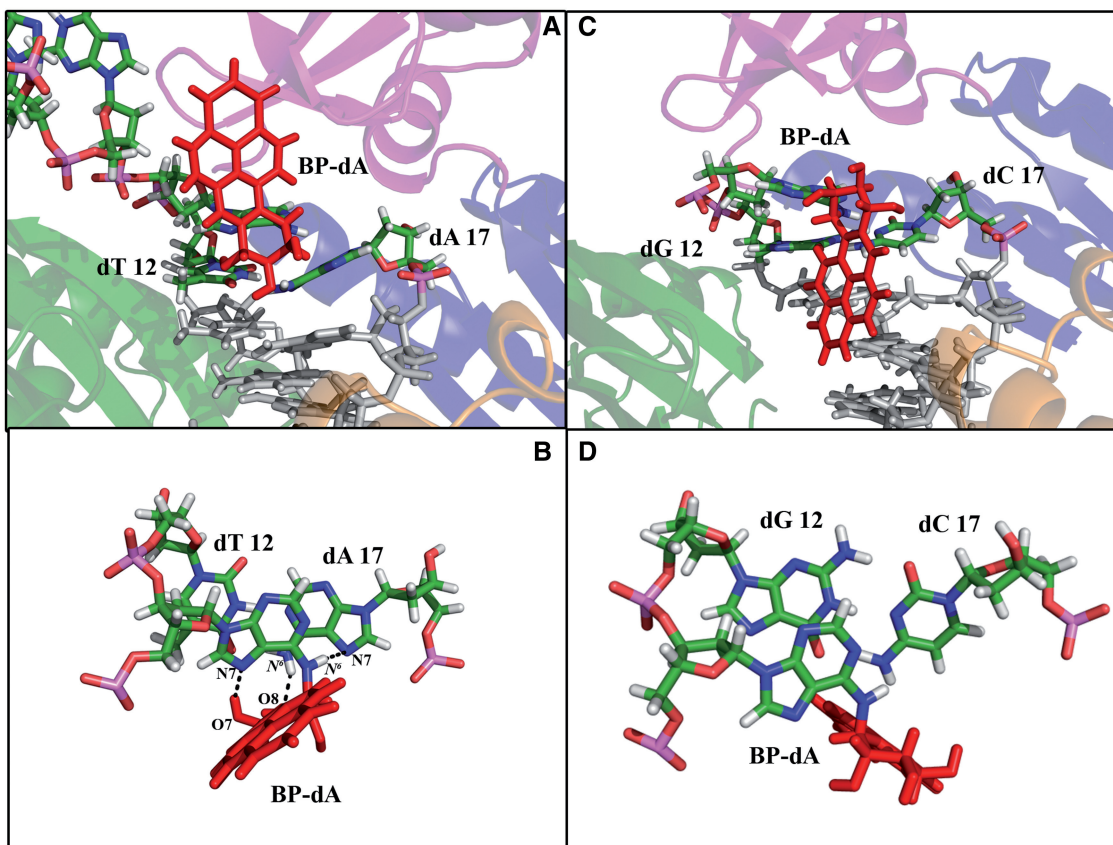
**Figure 1.** Adduct structure and DNA sequences studied. (A) Structure of the (+)-*trans-anti*-[BP]-*N*<sup>6</sup>-dA lesion (BP-dA).  $\chi$ : O4'-C1'-N9-C4;  $\alpha'$ : N1-C6-*N*<sup>6</sup>-C10;  $\beta'$ : C6-*N*<sup>6</sup>-C10-C9. (B) DNA sequences used in this study. A\* designates the BP-dA lesion.

### Binary complex unmodified control simulations show the expected *anti* conformation

The binary complex simulations with an unmodified *anti* templating dA showed normal active site organization. The templating base stacks very well with base 12 (Figure 1B), with average estimated Van der Waals stacking energies of  $-12.03$  kcal/mol for the 5'-GA\*TT-3' sequence (SeqI) and  $-12.83$  kcal/mol for the 5'-TA\*GA-3' sequence (SeqII) (Supplementary Figures S1 and S2). The lower energy of SeqII is due to the difference in the identity of base 12: dT in SeqI, and dG in SeqII. The purine-purine stacking interaction has more overlapping surface area, and thus lower energy. The primer terminus is correctly paired with base 12 in both simulations, and the simulations are very stable, showing little deviation over the course of the trajectory (Supplementary Movies S1 and S2, Supplementary Figures S1 and S2). The templating base is in the correct position and orientation to hydrogen bond with an incoming dTTP in both sequences.

### In binary complex BP-dA simulations with *anti* templating base, both blocking and nonblocking BP conformations are observed for each sequence

We performed 20 ns simulations on four BP-dA damaged binary complex models, examining two different initial  $\alpha'$ ,  $\beta'$  combinations for the lesion in each sequence (Figure 1B, Supplementary Table S1). In these structures, the templating base was placed in the *anti* domain of  $\chi$  in accordance with the binary crystal structures of polt (15).



**Figure 2.** Binary BP-dA active site structures. The most representative structures from the last cluster of each trajectory have been selected for illustrative purposes (see ‘Materials and Methods’ section for details). Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green;  $Mg^{2+}$ , purple. The nascent base pair and previously incorporated base pair are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, the other DNA is gray. The base numbering is as shown in Figure 1B. Bases 14–16 in the single-stranded overhang are not shown for clarity. Figure 2A and B are derived from a trajectory snapshot selected from snapshots within 100 frames of the most representative structure of the last cluster. This was done to better illustrate important active site features not clearly displayed in the most representative structure of the last cluster. (A) SeqI binary complex active site, major groove view. The binary complex is well ordered, with normal template base stacking and position. The BP rings are oriented in the 5' direction of the modified strand. A stereo version can be found in Supplementary Figure S25, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S3. (B) SeqI binary complex active site, top view. The templating base is intrahelical and stacks well with dT-12 and dA-17. Note the following hydrogen bonds that stabilize the 5' orientation of the BP rings: BP-dA O7H...BP-dA N7; BP-dA O8H...dA-17 N<sup>6</sup>; BP-dA N<sup>6</sup>...dA-17 N7. (C) SeqII binary complex active site, major groove view. The BP rings are oriented 3' along the modified strand, stacking against the major groove face of the DNA. This pulls the templating base toward the major groove, disrupting base stacking and opening the active site. A stereo version can be found in Supplementary Figure S26, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S4. (D) SeqII binary complex active site, top view. The templating base is displaced toward the major groove side of the nascent double helix, creating a more spacious active site which would decrease the barrier to rotation around the glycosidic bond. No hydrogen bonds are formed between the BP-dA and the surrounding sequence.

In our simulations in both sequences, one of the initial  $\alpha'$ ,  $\beta'$  combinations led to a blocking or otherwise non-productive structure, while the other resulted in a structure that would potentially allow for binding and incorporation of an incoming dNTP. The blocking structure in the 5'-GA\*TT-3' sequence (SeqI) places the BP lesion immediately 3' to the primer terminus, such that it is placed in the space that an incoming dNTP would occupy (Supplementary Figure S23). The non-productive structure in the 5'-TA\*GA-3' sequence (SeqII) shows the primer terminus displaced from the active site in the 5' direction of the primer strand, such that an incoming dNTP in its normal position would be unable to react with the terminal O3' (Supplementary Figure S24).

#### In binary complex BP-dA simulations with SeqI the templating BP-dA is stabilized in the *anti* glycosidic torsion domain by several hydrogen bonds

The potentially productive 5'-GA\*TT-3' sequence (SeqI) trajectory is stable, with little variation in the active site over the course of the trajectory (Supplementary Figure S3, Supplementary Movie S3). We observed a hydrogen bond between O8 on the BP adduct and N<sup>6</sup> on dA-18 (Figure 1B) for the first 10 ns of the trajectory that shifts to N<sup>6</sup> on dA-17 for the last 10 ns. We also observed a hydrogen bond between O7H on the adduct and N7 on the templating base that persists for the entire trajectory (Figure 2A and B, Supplementary Figure S3). In addition, there is a sporadic hydrogen bond between Gln59 and N3

on the templating base throughout the last 10 ns of the trajectory, as well as a sequence-specific hydrogen bond between N<sup>6</sup> of BP-dA and N7 of dA-17, the primer terminus (Supplementary Figure S3). These hydrogen bonds serve to maintain the templating BP-dA in the *anti* conformation, and also keep the bulk of the adduct from occupying the incoming dNTP binding site (Supplementary Movie S3). This structure suggests that in SeqI, an incoming dNTP would be presented with an *anti* templating BP-dA. Moreover, the hydrogen bonds that secure the adduct in place would increase the energy barrier between *anti* and *syn* conformations of the templating BP-dA. Our previous modeling and MD simulations (27) have shown that polt can tolerate a Watson–Crick pair in the active site if the alternative requires serious enzyme or DNA distortion, which would entail an excessive expenditure of energy.

#### **The SeqII BP-dA binary complex shows the templating *anti* BP-dA shifted toward the major groove**

The potentially productive 5'-TA\*GA-3' sequence (SeqII) structure with *anti* templating base shows a moderately flexible templating BP-dA, with the BP rings of the adduct finding a stable position on the major groove side of the nascent duplex after ~11 ns (Supplementary Movie S4, Supplementary Figure S4). No hydrogen bonds form between the BP-dA and the primer DNA strand in SeqII because there is no equivalent to the major groove side hydrogen bond acceptor dA-17 N7 in SeqI (Supplementary Figure S4). Instead, the BP rings extend into the major groove and pack tightly against the major groove face of the DNA duplex region for the latter half of the trajectory. This pulls the templating BP-dA toward the major groove, weakening the stacking interactions between the templating BP-dA and base 13 (Figures 1B, 2C and D, Supplementary Movie S4). The sugar on the templating base remains in the hydrophobic pocket formed by Gln59, Lys60 and Leu62. This displacement of the base toward the exterior of the helix would more easily allow rotation around the glycosidic bond, facilitating a transition to the *syn* domain. In addition, the displacement of the templating base creates a more open dNTP binding pocket. This more open active site may relieve some restrictions that hinder dGTP entry in SeqI.

#### **Most ternary complex simulations show good P $\alpha$ -O3' geometry and Mg<sup>2+</sup> coordination**

All but one of our ternary complex simulations, including both undamaged controls and damaged BP-dA simulations, show reaction-ready active sites for the entire trajectory as determined by the following criteria based on quantum mechanical/molecular mechanical studies (79) and a high resolution DNA polymerase crystal structure (80): P $\alpha$ -O3' distance is between 3.0 and 3.5 Å (Supplementary Table S4), P $\alpha$ -O3' in-line attack angles are less than 160° (Supplementary Table S4), Mg<sup>2+</sup> to Mg<sup>2+</sup> distances are 3.8–4.1 Å (Supplementary Table S4), 11 of 12 electronegative atoms coordinating the two Mg<sup>2+</sup> ions are within 2.2 Å of the Mg<sup>2+</sup>, and only the 12th

coordinating atom is within 3.1 Å (Supplementary Table S5–S6). Polt is known to increase in fidelity with Mn<sup>2+</sup> as the catalytic ion (81), perhaps due to the weaker coordination requirements of Mn<sup>2+</sup> as opposed to Mg<sup>2+</sup>. This may explain the less than ideal coordination for one Mg<sup>2+</sup> ion observed in our simulations. The only simulation that does not fit these criteria is the unmodified control simulation with incoming dATP in SeqII, which has several of the atoms coordinating the Mg<sup>2+</sup> ions at distances greater than 3.1 Å. This is unsurprising; dATP has a low rate of incorporation in unmodified DNA, as would be expected with poor Mg<sup>2+</sup> coordination.

#### **Ternary complex unmodified control simulations with *syn* templating dA show well-organized active sites with incoming correct partner dTTP**

Eight control simulations with an unmodified *syn* templating dA were performed. In the primer extension experiments, there was no sequence effect observed with the unmodified controls; both sequences incorporated the correct partner dT, with minor incorporation of dG and dA, and dCTP was not measurably incorporated (41). Our simulations are in agreement with this result. In both sequences, the unmodified control simulations with incoming dTTP show two strong Hoogsteen hydrogen bonds, well-ordered active sites, coplanar bases, and the incoming dNTP stacked with the primer terminus (Supplementary Figures S5 and S6, Supplementary Movies S5 and S6). The trajectories are stable, and these features are maintained with normal minor fluctuations in the active site region throughout the trajectories (Supplementary Figures S5 and S6, Supplementary Movies S5 and S6).

#### **Ternary complex unmodified control simulations with *anti* dA exhibit shortened C1' to C1' distances and loss of hydrogen bonds**

We also performed simulations in each sequence with *anti* templating unmodified dA and incoming dTTP to examine the stability of a Watson–Crick AT pair in polt. Both of these simulations began with a well-aligned Watson–Crick AT pair with a standard C1' to C1' distance of 10.6 Å. However, during the simulation, the C1' to C1' distance shortened to 10.1 Å in the 5'-GA\*TT-3' sequence (SeqI) (Supplementary Figure S7) and 10.3 Å in the 5'-TA\*GA-3' sequence (SeqII) (Supplementary Figure S8) due to constraints of the polt active site. In SeqI, the decreased C1' to C1' distance causes the templating dA and incoming dTTP to buckle and become misaligned, losing both Watson–Crick hydrogen bonds at ~14 ns (Supplementary Figure S7, Supplementary Movie S7). In SeqII, the slightly longer C1' to C1' distance is enabled by the displacement of the three base overhang 5' to the templating dA (dT-14, dT-15 and dT-16) (Figure 1B) toward the major groove side of the polymerase (Supplementary Figure S8, Supplementary Movie S8). dT-16 stacks with Tyr355 in the little finger domain and forms a hydrogen bond with Arg357, and dT-14 moves toward the major groove. This arrangement pulls the sugar of the templating dA away from the

polymerase, moving the templating base toward the major groove and widening the C1' to C1' distance (Supplementary Figure S8, Supplementary Movie S8). Interestingly, there is a brief period of  $\sim 0.5$  ns in the latter part of the trajectory where the C1' to C1' distance decreases sharply to  $\sim 8.6$  Å (Supplementary Figure S8), forcing the templating dA into the major groove and losing hydrogen bonding similar to the conformation seen in SeqI. This decrease is brought about by the temporary resumption of the normal, stacked position of dT-14 and dT-15, though dT-16 remains stacked with Tyr355 in the little finger domain. However, the C1' to C1' distance lengthens as soon as the overhang moves back to the major groove side of the polymerase.

#### Mismatched ternary complex unmodified control simulations show disorganized active sites

The dCTP unmodified control simulations with *anti* templating base show no hydrogen bonds in either sequence, and also exhibit various active site defects, namely long C1' to C1' distances, poor stacking of incoming dNTP with the primer terminus, and oddly positioned templating bases (Supplementary Figures S9 and S10, Supplementary Movies S9 and S10). Both dCTP trajectories show the templating dA moving away from the incoming dCTP, ending in a position far from the active site (Supplementary Movies S9 and S10). In addition, the templating dA's sugar moves out of the hydrophobic pocket formed by Gln59, Lys60 and Leu62, indicating strain.

In the undamaged control incoming dATP simulations with templating dA *syn*, we observed different outcomes in each sequence. In the 5'-GA\*TT-3' sequence (SeqI), the dA-dA mispair shows a buckling of the nascent base pair, with the formation of one hydrogen bond between N<sup>6</sup> on the templating base and N1 on the incoming dATP (Supplementary Movie S11, Supplementary Figure S11). In the 5'-TA\*GA-3' sequence (SeqII), the templating dA is completely displaced from the active site, and the incoming dATP is stacked poorly with primer terminus (Supplementary Movie S12, Supplementary Figure S12). These two poorly organized active sites both support the experimental results showing little dATP incorporation in the unmodified case (41).

Unmodified control simulations with incoming dGTP and templating dA *syn* also show two strained active site conformations. In SeqI, the incoming dGTP forms two hydrogen bonds with the templating dA: one is between dA N<sup>6</sup> and dGTP O<sup>6</sup>, while the other is a bifurcated hydrogen bond where dA N7 and dA N1 hydrogen bond with dGTP N<sup>2</sup> (Supplementary Figure S13). The nascent base pair buckles and twists in an attempt to minimize the C1'-C1' distance (Supplementary Movie S13). SeqII achieves the same hydrogen bonds, but occasionally the templating dA moves away from the active site, similar to the behavior exhibited with incoming dATP in SeqII (Supplementary Movie S14, Supplementary Figure S14). These results support the poor dGTP incorporation observed experimentally in much the same fashion as the dATP simulations above.

Overall, the unmodified control simulations with templating dA *syn* show that of the four incoming dNTPs, only dTTP can form a stable, well-organized active site. Incoming dATP and dGTP exhibit defects in both sequences, helping to explain the similarly weak incorporation profile observed in primer extension experiments (41). Incoming dCTP simulations show very disorganized active sites, supporting the lack of dCTP incorporation in the primer extension experiments (41).

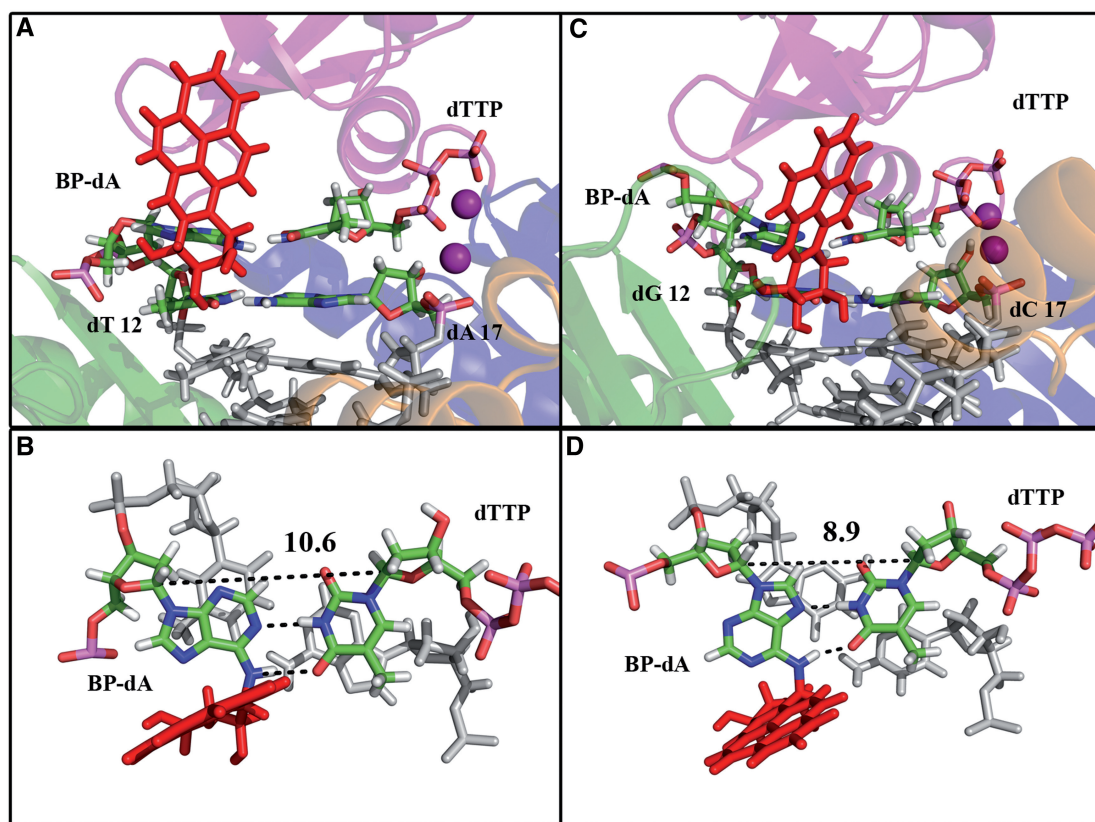
#### Ternary complex BP-dA simulations of both sequences show well-organized active sites with incoming dTTP utilizing Watson-Crick base pairing in SeqI and Hoogsteen base pairing in SeqII

The BP-dA ternary complex simulation with incoming dTTP in the 5'-GA\*TT-3' sequence (SeqI) shows an *anti* template and an *anti* incoming nucleotide, forming a classical Watson-Crick pair (Figure 3A and B). Both Watson-Crick hydrogen bonds are present and strong, and the active site is well organized, with good base stacking (Figure 3A and B, Supplementary Figure S15). The trajectory is very stable, with only subtle fluctuations in the active site geometry (Supplementary Movie S15). In the 5'-TA\*GA-3' sequence (SeqII) with *syn* templating BP-dA, the incoming dTTP simulation performed well, with a normal Hoogsteen dA-dT pair with two hydrogen bonds (Figure 3C and D, Supplementary Figure S16). Base stacking of the damaged template is somewhat disrupted, with the templating BP-dA extending toward the major groove, similar to the position observed with the SeqII binary BP-dA complex (Figure 3C and D, Supplementary Figure S16, Supplementary Movie S16). The SeqII trajectory is also fairly stable, with few changes in active site organization throughout the trajectory (Supplementary Movie S16). In both sequences, incoming dTTP forms the active site with the strongest hydrogen bonds and best overall organization among the four incoming dNTPs, as is to be expected from the predominantly error-free bypass seen in the experimental data (41) (Figure 3, Supplementary Figures S15-S16).

#### Ternary complex BP-dA simulations of both sequences show disorganized active sites with incoming dCTP

Both simulations with incoming dCTP displayed distorted active sites, though the bulky adduct seems to prevent the more extreme displacement of the templating BP-dA toward the little finger domain and very wide C1'-C1' distances observed in the unmodified control simulations (Supplementary Figures S9, S10, S17, S18). No base pairing is observed between the templating BP-dA and the incoming dCTP in either case, and the base pair geometry is distorted in both sequences, leading to long C1'-C1' distances (Supplementary Figures S17-S18). Both simulations show a flexible templating base, with the SeqII slightly more mobile than SeqI (Supplementary Movies S17-S18). The poor performance of the incoming dCTP simulations is unsurprising given the unstable nature of dC-dA mismatches (82) as well as the low proportion of dCTP incorporation that is observed in both sequences (41).





**Figure 3.** Incoming dTTP BP-dA ternary structure active sites. The most representative structures from the last cluster of each trajectory have been selected for illustrative purposes (see ‘Materials and Methods’ section for details). Insets show overhead views of the nascent base pair and the pair formed by base 12 and base 17 (numbering shown in Figure 1B). Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green;  $Mg^{2+}$ , purple. The nascent base pair, the pair formed by base 12 and base 17 and the 5' overhang are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is gray. (A) Sequence I incoming dTTP ternary structure active site, major groove view. The incoming dTTP forms a normal Watson-Crick base pair with the templating BP-dA, with two hydrogen bonds and a C1'-C1' distance of 10.6 Å. A stereo version can be found in Supplementary Figure S27, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S15. (B) Sequence I incoming dTTP ternary structure active site, top view. (C) Sequence II incoming dTTP ternary structure active site, major groove view. The incoming dTTP forms a normal Hoogsteen base pair with the templating BP-dA. The active site is well organized, with two Hoogsteen hydrogen bonds and a C1'-C1' distance of 8.9 Å. A stereo version can be found in Supplementary Figure S28, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S16. (D) Sequence II incoming dTTP ternary structure active site, top view.

### Ternary complex BP-dA simulations show strained active sites for incoming dATP in both sequences

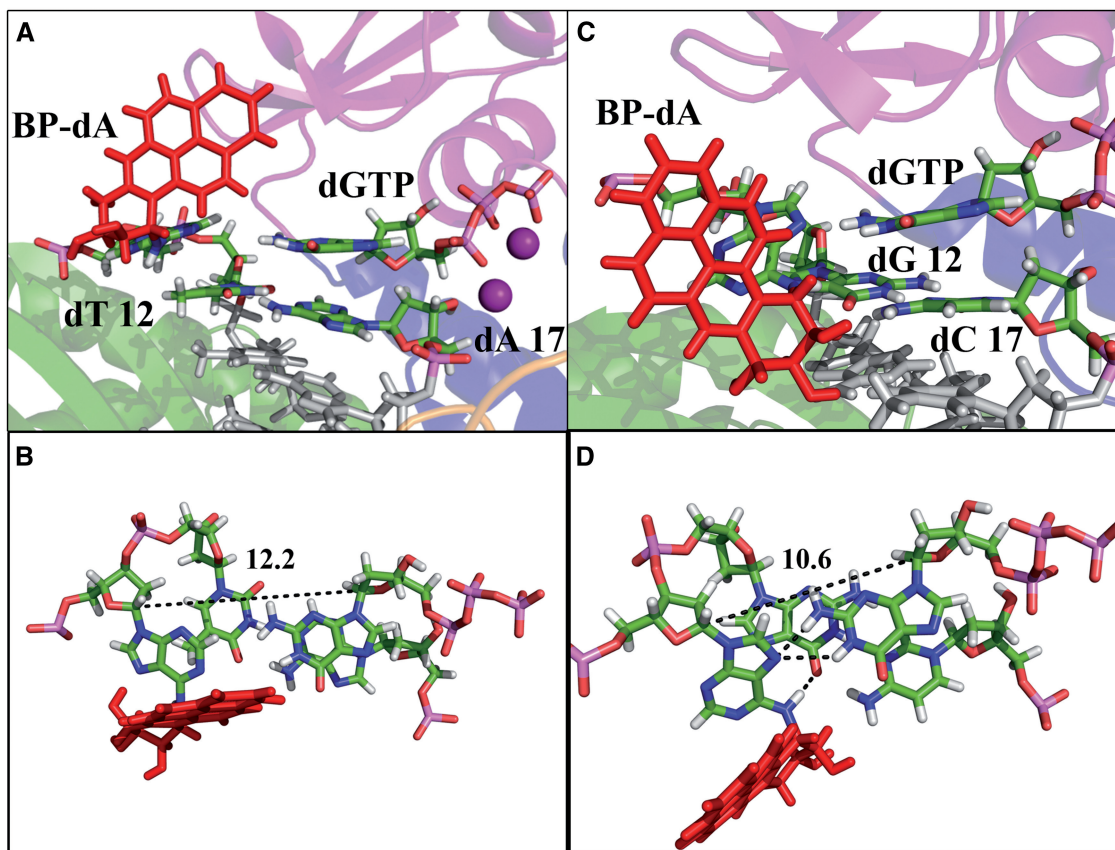
The BP-dA simulation with incoming dA in the 5'-GA\*TT-3' sequence (SeqI) shows an active site very similar to the SeqI unmodified dATP control simulation described above (Supplementary Movie S11, Supplementary Figure S11), with a BP-dA  $N^6$  to dATP N1 hydrogen bond and a buckled and twisted nascent base pair (Supplementary Movie S19, Supplementary Figure S19). This strained active site is consistent with the primer extension data which shows low levels of dA incorporation opposite both SeqI BP-dA and undamaged template. In the 5'-TA\*GA-3' sequence (SeqII), our simulations show the BP-dA moving toward the major groove and twisting, with no hydrogen bonding between the template and incoming dATP, similar to the SeqII binary complex (Figure 2C and D, Supplementary Movie S19, Supplementary Figure S19).

In both SeqI and SeqII dATP simulations, the sugar of the templating base is held in the hydrophobic pocket

formed by Gln59, Lys60 and Leu62. This causes distortions in the mismatched pair in order to attain a C1'-C1' distance of  $\sim 10$  Å, which can be accommodated within polt according to MD simulations (27) (Supplementary Figures S19, S20). In both cases, the templating BP-dA moves toward the major groove and twists in order to allow the incoming dNTP to occupy the active site (Supplementary Movie S19, S20). These strained conformations support the relatively low rate of dA misincorporation seen in the primer extension experiments (41).

### Ternary complex BP-dA simulations with incoming dGTP show differences in hydrogen bonding between SeqI and SeqII that may explain an observed enhanced efficiency of dGTP incorporation in SeqII

The SeqI and SeqII simulations with incoming dGTP may help explain the  $\sim 14$ -fold greater dGTP misincorporation efficiency in SeqII as compared to SeqI, based on the ratios of their  $V_{max}/K_m$  values (41). In the 5'-GA\*TT-3'



**Figure 4.** Incoming dGTP BP-dA ternary structure active sites. The most representative structures from the last cluster of each trajectory have been selected for illustrative purposes (see ‘Materials and Methods’ section for details). Insets show overhead views of the nascent base pair and the pair formed by base 12 and base 17 (numbering shown in Figure 1B). Color code: Fingers domain, magenta; palm domain, blue; thumb domain, orange; little finger domain, green;  $Mg^{2+}$ , purple. The nascent base pair, the pair formed by base 12 and base 17, and the 5' overhang are colored by atom: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; phosphorus, magenta. The BP lesion is red, other DNA is gray. (A) Sequence I incoming dGTP ternary complex structure active site, major groove view. The bulk of the incoming dGTP causes the templating BP-dA to move out of the active site, with an average C1'–C1' distance of 12.2 Å. This increased distance results in the loss of hydrophobic contacts between the sugar of BP-dA and protein residues Gln59, Lys60 and Leu62. In addition, there are no hydrogen bonds between the incoming dGTP and BP-dA. A stereo version can be found in Supplementary Figure S29, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S21. (B) Sequence I incoming dGTP ternary complex structure active site, top view. (C) Sequence II incoming dGTP ternary complex structure active site, major groove view. The templating BP-dA twists to accommodate the incoming dGTP while maintaining a relatively short C1'–C1' distance of 10.6 Å. The sugar of the templating base is held by residues Gln59, Lys60 and Leu62. One bifurcated hydrogen bond is present between N1 on the incoming dGTP and  $N^6$  and N7 on the BP-dA, with a second hydrogen bond between BP-dA  $N^6$  and dG-12  $O^6$ . A stereo version can be found in Supplementary Figure S30, and a movie of the most representative structures from each cluster can be seen in Supplementary Movie S22. (D) Sequence II incoming dGTP ternary complex structure active site, top view.

sequence (SeqI) simulation with incoming *syn* dGTP and *anti* BP-dA, the purine–purine mismatch cannot be accommodated in the narrow active site of polt, resulting in the displacement of the templating BP-dA away from the active site and toward the little finger domain (Supplementary Movie S21). The C1'–C1' distance increases to ~12 Å, and the sugar on the templating BP-dA disengages from the hydrophobic pocket that normally holds it (Figure 4A and B). There are no hydrogen bonds between the incoming dGTP and the templating BP-dA (Figure 4A and B, Supplementary Figure S21). This disorganized active site shows considerable strain, supporting the observed dearth of dGTP incorporation in SeqI (41).

In the 5'-TA\*GA-3' sequence (SeqII), the *syn* BP-dA is very flexible, with early conformations resembling

the incoming dGTP SeqII unmodified controls (Supplementary Figures S13–S14). As the simulation progresses, the templating BP-dA shifts toward the major groove and twists, forming a bifurcated hydrogen bond between N1 and  $N^2$  on the *anti* dGTP and N7 on the templating BP-dA. BP-dA  $N^6$  forms a second hydrogen bond with dGTP  $O^6$ , though it later shifts to  $O^6$  on dG-12 (Figures 1B, 4C and D). The C1'–C1' distance is ~10.5 Å, and the sugar of the templating BP-dA is held in place by residues Gln59, Lys60 and Leu62 (Figure 4C and D, Supplementary Movie S22, Supplementary Figure S22). Thus the active site in SeqII exhibits better organization than in SeqI, supporting the increased relative efficiency of dGTP incorporation.

It should be noted that in order to accommodate a wider C1'–C1' distance in the Watson–Crick BP-dA/

dTTP pair as well as in the purine/purine mismatches, residues Gln59, Lys60 and Leu62 move  $\sim 2\text{\AA}$  relative to the remainder of the fingers domain and the active site. The protein-active site is otherwise essentially unchanged in these structures.

## DISCUSSION

We performed binary complex simulations containing only DNA and the enzyme with the templating dA *anti* in order to better understand the structure of the active site without an incoming dNTP. The templating base was in the *anti* conformation in all binary complex simulations in accordance with the published binary complex crystal structures (15). Both sequences exhibited one blocking or nonpermissive conformation, and one that would allow entrance of a dNTP. In the permissive structures, hydrogen bonding partners available to the hydroxyl residues on the BP adduct resulted in different conformations for the templating base. In the 5'-GA\*TT-3' sequence (SeqI) bases 17 and 18 on the primer strand are dA (Figure 1B). This presents the adduct with the opportunity to hydrogen bond to  $N^6$  on these primer terminal residues. This in turn positions the adduct so that one of the hydroxyl groups can hydrogen bond to N7 on the templating base. Ne2 on Gln59 also hydrogen bonds with N3 on BP-dA for the latter half of the trajectory (Supplementary Figure S3). Together, these hydrogen bonds stabilize an *anti* template conformation, leading to correct selection of dTTP via a Watson-Crick base pair. In SeqII, however, the primer terminal residues are dC and dT, both of which fail to present an accessible hydrogen bonding partner for the BP hydroxyl groups (Figure 1B). Therefore, in the 5'-TA\*GA-3' sequence (SeqII), BP-dA is more flexible and can assume a different conformation, which pulls the templating dA toward the major groove in order to maximize van der Waals contacts between the BP rings system and the nascent double helix (Figure 1B and D, Supplementary Figures S4 and S26). In this position, the BP-dA would be able to rotate from *anti* to *syn* upon entry of the dNTP with relative ease, leading to the predominant correct dTTP selection via a Hoogsteen base pair, similar to unmodified DNA.

While both sequences correctly incorporate dT in favor of mismatched nucleotides, SeqII incorporates dG with  $\sim 14$ -fold higher efficiency than SeqI (Table 1) (41). Our simulations suggest that this difference could be due to the glycosidic bond conformation favored in each sequence: *anti* in the 5'-GA\*TT-3' sequence (SeqI) and *syn* in the 5'-TA\*GA-3' sequence (SeqII). The *syn* BP-dA is able to form a bifurcated hydrogen bond between N1 and  $N^2$  on

dGTP and N7 on the templating BP-dA as well as a second hydrogen bond between BP-dA  $N^6$  and  $O^6$  on the incoming dGTP or base dG-12 (Figures 1B, 4C and D). *Anti* BP-dA cannot form any hydrogen bond with incoming dGTP (Figure 4A and B, Supplementary Figure S21), suggesting a rationale for the relatively decreased efficiency of dGTP incorporation in SeqI. In addition, the *anti* template conformation in SeqI leads to longer C1'-C1' distances, which are known to be disfavored in polt (14).

Our data shows that polt can tolerate both Hoogsteen and Watson-Crick compatible C1'-C1' distances in the active site, but with a preference for shorter distances. However, C1'-C1' distances beyond  $\sim 10.6\text{\AA}$  are more strongly disfavored. The preferred shorter C1'-C1' distance qualitatively explains several observed differences in incorporation efficiency among the different sequences and incoming dNTPs (Table 1). For example, the efficiency ( $V_{\max}/K_m$ ) of dGTP incorporation opposite *syn* BP-dA in SeqII (Table 1), with a C1'-C1' distance of  $10.1\text{\AA}$  (Supplementary Figure S20), is  $\sim 14$ -fold greater than that of dGTP incorporation opposite *anti* BP-dA in SeqI, with a C1'-C1' distance of  $12.2\text{\AA}$  (Supplementary Figure S19). This preference for shorter C1'-C1' distances can also explain the  $\sim 1.7$ -fold greater efficiency for incorporation of dTTP opposite the BP-dA lesion via a Hoogsteen base pair ( $\sim 8.9\text{\AA}$  C1'-C1' distance) in SeqII compared to SeqI, where dTTP is incorporated via a Watson-Crick pair ( $\sim 10.6\text{\AA}$  C1'-C1' distance) (Table 1, Figure 3 and Supplementary Figures S15, S16). Another example is the case of dATP incorporation opposite *anti* BP-dA in SeqI as compared with dGTP incorporation in the same sequence: here, the incorporation efficiency for dATP is  $\sim 9$ -fold greater than for dGTP (Table 1), in line with the respective C1'-C1' distances of  $9.8$  and  $12.2\text{\AA}$  (Supplementary Figures S19 and S21). Also, the larger  $12.2\text{\AA}$  C1'-C1' distance allows only one hydrogen bond with partial occupancy between *anti* BP-dA and incoming dGTP (Supplementary Figure S21), while with the  $9.8\text{\AA}$  distance in the dATP simulation one high-occupancy hydrogen bond forms between the *anti* BP-dA and dATP (Supplementary Figure S19). In SeqII, the similar efficiency of dATP and dGTP ( $\sim 1.8$ -fold difference in efficiency) (Table 1) is in concert with their similar C1'-C1' distances of  $9.5\text{\AA}$  for dATP (Supplementary Figure S20) and  $8.2\text{\AA}$  for dGTP (Supplementary Figure S22), both opposite *syn* BP-dA.

Our results from the *anti* templating unmodified dA with incoming dTTP simulations are also interesting. Both simulations show a shortened C1'-C1' distance (Supplementary Figures S7-S8), showing agreement with prior studies indicating that polt does not favor a Watson-Crick pair with undamaged DNA (14-16,25,26). In SeqI, the Watson-Crick AT pair buckles under the strain, losing both hydrogen bonds (Supplementary Movie S7). In SeqII, the templating dA is pulled toward the major groove by a movement of the bases 5' to the templating dA, allowing for a slightly longer C1'-C1' distance ( $10.3\text{\AA}$ ) than observed in SeqI ( $10.1\text{\AA}$ ) (Supplementary Figures S7-S8, Supplementary Movie S8). This adaptation is

**Table 1.**  $V_{\max}/K_m$  values for incorporation of all four nucleotides opposite BP-dA in SeqI and SeqII

	dATP	dCTP	dGTP	dTTP
SeqI (5'-GA*TT-3')	0.011	0.001	0.0012	1.21
SeqII (5'-TA*GA-3')	0.03	0.0014	0.017	2.04

Source: Ref. (41). The units of  $V_{\max}/K_m$  are the percentage of primer extension product/min/ $\mu\text{mol}$  nucleotide.

similar to that observed in our previous study with a Watson–Crick unmodified GC pair in polt (27).

The biological function of human DNA polymerase  $\iota$  has been something of a mystery since its discovery and characterization over a decade ago (13). Polt knockout mice show no unusual phenotype, though they are more susceptible to urethane-induced mutations than control mice (83). Mice deficient in polt also appear prone to formation of mesenchymal tumors (84). Polt is capable of bypassing a wide variety of DNA lesions in *in vitro* studies, but which lesions are bypassed by polt in animals has been less certain. However, evidence has been accumulating about which strongly supports the role of polt in bypass of thymine dimers in an error-prone fashion (84,85), notably in human XPV-derived cells, which lack functional polymerase  $\eta$  (86,87). Whether polt is involved in the bypass of bulky lesions in humans, such as those derived from the BP metabolites examined in this study, remains to be determined.

Our application of a quality threshold clustering algorithm aids in understanding the dynamic structural properties of long trajectories; and its employment in clustering MD trajectories is new, as far as we are aware. As computing power increases, MD simulations are being performed on longer and longer time scales, and more initial structures are created and examined. This leads to a superabundance of data, and necessitates novel approaches for organizing and analyzing output. Looking for structural motifs in a trajectory of 20 000 + frames by eye is a daunting and error-prone task. Therefore, we have adapted and implemented an algorithm for clustering MD trajectories in order to understand the dynamic structural properties of the system, the details and rationale of which are given in the Supplementary Data. We clustered our trajectories and selected the most representative frame from each cluster, i.e. the member of a cluster with the lowest active site RMSD to all other cluster members. These most representative structures allow us to facilely achieve an understanding of the substates sampled in our simulations. Animations derived from the most representative structures of each cluster are demonstrated in Supplementary Movies S1–S22.

Stable structures often exhibit only normal thermal motions over the trajectory, but the clustering approach can facilitate the discovery of rarely sampled substates or periodic motions in the system that could otherwise be lost. An example from the current work is the unmodified control simulation in the 5'-TA\*GA-3' sequence (SeqII) with incoming dGTP. In the incoming dGTP simulation, the templating base is often aligned with the incoming dGTP, but on occasion it leaves the active site and moves toward the little finger domain (Supplementary Movie S14). This fluctuation between the two positions would be masked by examining only the latter portion of the trajectory or ensemble average data values.

The application of this clustering algorithm to MD simulations generates a wealth of additional data about the trajectories analyzed. While development of techniques to take advantage of this information is beyond the scope of this article, there are a number of possible ways

to potentially gain insight by the analysis of clustering metadata. For example, the number of clusters returned by our algorithm may indicate the inherent barriers to transition between different substates, or the percentage of frames not in a cluster may indicate the flexibility of a given protein or complex. If cluster metadata were combined with other energetic analysis techniques, one could access the relationship of cluster size and membership to relative free energy, enthalpy and entropy (88). Future analysis using this clustering method has potential to give deeper insight into the energetics and dynamics of long MD trajectories.

## CONCLUSION

Our MD simulations are in agreement with the observation (41) that human DNA polymerase  $\iota$  can incorporate dT opposite a BP-dA adduct in a relatively error-free manner. We find that the local sequence context can, through specific hydrogen bonds, influence the glycosidic bond conformation of the templating base, leading to dT incorporation via a Watson–Crick base pair in a 5'-CAGA\*TT-3' sequence context (SeqI) and via a Hoogsteen base pair in a 5'-TTTA\*GA-3' sequence context (SeqII). Moreover, this same sequence-specific effect can offer a structural rationale for the greater efficiency of mutagenic dGTP incorporation in SeqII as compared to SeqI. Polt's versatility implies that this enzyme can adjust to local structural conditions in order to bypass bulky lesions by strategies that utilize Watson–Crick, Hoogsteen and perhaps other base pairing possibilities (17). There is an emerging accumulation of data indicating that polt can act as an error-prone backup to other bypass polymerases, notably pol $\eta$  (86,87). Polt's structural versatility may be related to such a backup role, providing it with the capability to bypass various lesions, albeit with mutagenic consequences in some cases.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Dr Olga Rechkoblit and Prof. Robert Shapiro for their very helpful advice.

## FUNDING

National Institutes of Health [CA28038 to S.B.] and the National Science Foundation through TeraGrid resources provided by the Texas Advanced Computing Center. Support for computational infrastructure and systems management was also provided by National Institutes of Health [CA75449 to S.B.]. Funding for open access charges: National Institutes of Health [CA28038 to S.B.]

*Conflict of interest statement.* None declared.

## REFERENCES

- Luch, A. (2005) Nature and nurture—lessons from chemical carcinogenesis. *Nat. Rev. Cancer*, **5**, 113–125.
- Clapp, R.W., Jacobs, M.M. and Loechler, E.L. (2008) Environmental and occupational causes of cancer: new evidence 2005–2007. *Rev. Environ. Health*, **23**, 1–37.
- Friedberg, E.C., Lehmann, A.R. and Fuchs, R.P. (2005) Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol. Cell*, **18**, 499–505.
- Lehmann, A.R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J.F., Kannouche, P.L. and Green, C.M. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst)*, **6**, 891–899.
- Waters, L.S., Minesinger, B.K., Wiltout, M.E., D'Souza, S., Woodruff, R.V. and Walker, G.C. (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.*, **73**, 134–154.
- Yang, W. and Woodgate, R. (2007) What a difference a decade makes: insights into translesion DNA synthesis. *Proc. Natl Acad. Sci. USA*, **104**, 15591–15598.
- McDonald, J.P., Rapic-Otrin, V., Epstein, J.A., Broughton, B.C., Wang, X., Lehmann, A.R., Wolgemuth, D.J. and Woodgate, R. (1999) Novel human and mouse homologs of *Saccharomyces cerevisiae* DNA polymerase  $\eta$ . *Genomics*, **60**, 20–30.
- Tissier, A., McDonald, J.P., Frank, E.G. and Woodgate, R. (2000) polt, a remarkably error-prone human DNA polymerase. *Genes Dev.*, **14**, 1642–1650.
- Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T. et al. (2001) The Y-family of DNA polymerases. *Mol. Cell*, **8**, 7–8.
- Zhang, Y., Yuan, F., Wu, X. and Wang, Z. (2000) Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase  $\iota$ . *Mol. Cell Biol.*, **20**, 7099–7108.
- McCulloch, S.D. and Kunkel, T.A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.*, **18**, 148–161.
- Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S. and Aggarwal, A.K. (2000) Eukaryotic polymerases  $\iota$  and  $\zeta$  act sequentially to bypass DNA lesions. *Nature*, **406**, 1015–1019.
- Vidal, A.E. and Woodgate, R. (2009) Insights into the cellular role of enigmatic DNA polymerase  $\iota$ . *DNA Repair (Amst)*, **8**, 420–423.
- Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L. and Aggarwal, A.K. (2004) Replication by human DNA polymerase- $\iota$  occurs by Hoogsteen base-pairing. *Nature*, **430**, 377–380.
- Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2006) An incoming nucleotide imposes an *anti* to *syn* conformational change on the templating purine in the human DNA polymerase- $\iota$  active site. *Structure*, **14**, 749–755.
- Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2005) Human DNA polymerase  $\iota$  incorporates dCTP opposite template G via a G.C<sup>+</sup> Hoogsteen base pair. *Structure*, **13**, 1569–1577.
- Kirouac, K.N. and Ling, H. (2009) Structural basis of error-prone replication and stalling at a thymine base by human DNA polymerase  $\iota$ . *EMBO J.*, **28**, 1644–1654.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer, New York.
- Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2009) DNA synthesis across an abasic lesion by human DNA polymerase  $\iota$ . *Structure*, **17**, 530–537.
- Rechkoblit, O., Zhang, Y., Guo, D., Wang, Z., Amin, S., Krzeminsky, J., Louneva, N. and Geacintov, N.E. (2002) trans-Lesion synthesis past bulky benzo[a]pyrene diol epoxide N<sup>2</sup>-dG and N<sup>6</sup>-dA lesions catalyzed by DNA bypass polymerases. *J. Biol. Chem.*, **277**, 30488–30494.
- Washington, M.T., Minko, I.G., Johnson, R.E., Wolffe, W.T., Harris, T.M., Lloyd, R.S., Prakash, S. and Prakash, L. (2004) Efficient and error-free replication past a minor-groove DNA adduct by the sequential action of human DNA polymerases  $\iota$  and  $\kappa$ . *Mol. Cell Biol.*, **24**, 5687–5693.
- Wolffe, W.T., Johnson, R.E., Minko, I.G., Lloyd, R.S., Prakash, S. and Prakash, L. (2005) Human DNA polymerase  $\iota$  promotes replication through a ring-closed minor-groove adduct that adopts a *syn* conformation in DNA. *Mol. Cell Biol.*, **25**, 8748–8754.
- Choi, J.Y. and Guengerich, F.P. (2006) Kinetic evidence for inefficient and error-prone bypass across bulky N<sup>2</sup>-guanine DNA adducts by human DNA polymerase  $\iota$ . *J. Biol. Chem.*, **281**, 12315–12324.
- Pence, M.G., Blans, P., Zink, C.N., Hollis, T., Fishbein, J.C. and Perrino, F.W. (2009) Lesion bypass of N<sup>2</sup>-ethylguanine by human DNA polymerase  $\iota$ . *J. Biol. Chem.*, **284**, 1732–1740.
- Johnson, R.E., Haracska, L., Prakash, L. and Prakash, S. (2006) Role of Hoogsteen edge hydrogen bonding at template purines in nucleotide incorporation by human DNA polymerase  $\iota$ . *Mol. Cell Biol.*, **26**, 6435–6441.
- Johnson, R.E., Prakash, L. and Prakash, S. (2005) Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase  $\iota$ . *Proc. Natl Acad. Sci. USA*, **102**, 10466–10471.
- Donny-Clark, K., Shapiro, R. and Broyde, S. (2009) Accommodation of an N-(deoxyguanosin-8-yl)-2-acetylaminofluorene adduct in the active site of human DNA polymerase  $\iota$ : Hoogsteen or Watson–Crick base pairing? *Biochemistry*, **48**, 7–18.
- Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. and Aggarwal, A.K. (2006) Hoogsteen base pair formation promotes synthesis opposite the 1,N<sup>6</sup>-ethenodeoxyadenosine lesion by human DNA polymerase  $\iota$ . *Nat. Struct. Mol. Biol.*, **13**, 619–625.
- Johnson, R.E., Yu, S.L., Prakash, S. and Prakash, L. (2007) A role for yeast and human translesion synthesis DNA polymerases in promoting replication through 3-methyl adenine. *Mol. Cell Biol.*, **27**, 7198–7205.
- Posky, B.S., Frank, E.G., Berry, D.A., Vennall, G.P., McDonald, J.P. and Woodgate, R. (2008) Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog *in vitro* and methyl methanesulfonate-induced DNA damage *in vivo*. *Nucleic Acids Res.*, **36**, 2152–2162.
- Choi, J.Y., Lim, S., Eoff, R.L. and Guengerich, F.P. (2009) Kinetic analysis of base pairing preference for nucleotide incorporation opposite template pyrimidines by human DNA polymerase  $\iota$ . *J. Mol. Biol.*, **389**, 264–274.
- Conney, A.H. (1982) Induction of microsomal enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G. H. A. Clowes Memorial Lecture. *Cancer Res.*, **42**, 4875–4917.
- Burczynski, M.E., Harvey, R.G. and Penning, T.M. (1998) Expression and characterization of four recombinant human dihydrodiol dehydrogenase isoforms: oxidation of *trans*-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene to the activated *o*-quinone metabolite benzo[a]pyrene-7,8-dione. *Biochemistry*, **37**, 6781–6790.
- Cavalieri, E.L. and Rogan, E.G. (1995) Central role of radical cations in metabolic activation of polycyclic aromatic hydrocarbons. *Xenobiotica*, **25**, 677–688.
- Meehan, T. and Straub, K. (1979) Double-stranded DNA stereoselectively binds benzo[a]pyrene diol epoxides. *Nature*, **277**, 410–412.
- Cheng, S.C., Hilton, B.D., Roman, J.M. and Dipple, A. (1989) DNA adducts from carcinogenic and noncarcinogenic enantiomers of benzo[a]pyrene dihydrodiol epoxide. *Chem. Res. Toxicol.*, **2**, 334–340.
- Szeliga, J. and Dipple, A. (1998) DNA adduct formation by polycyclic aromatic hydrocarbon dihydrodiol epoxides. *Chem. Res. Toxicol.*, **11**, 1–11.
- Buening, M.K., Wislocki, P.G., Levin, W., Yagi, H., Thakker, D.R., Akagi, H., Koreeda, M., Jerina, D.M. and Conney, A.H. (1978) Tumorigenicity of the optical enantiomers of the diastereomeric benzo[a]pyrene 7,8-diol-9,10-epoxides in newborn mice: exceptional activity of (+)-7 $\beta$ ,8 $\alpha$ -dihydroxy-9 $\alpha$ ,10 $\alpha$ -epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene. *Proc. Natl Acad. Sci. USA*, **75**, 5358–5361.
- Page, J.E., Zajc, B., Oh-hara, T., Lakshman, M.K., Sayer, J.M., Jerina, D.M. and Dipple, A. (1998) Sequence context profoundly influences the mutagenic potency of *trans*-opened benzo[a]pyrene 7,8-diol 9,10-epoxide-purine nucleoside adducts in site-specific mutation studies. *Biochemistry*, **37**, 9127–9137.
- Chary, P., Latham, G.J., Roberson, D.L., Kim, S.J., Han, S., Harris, C.M., Harris, T.M. and Lloyd, R.S. (1995) *In vivo* and *in vitro* replication consequences of stereoisomeric

- benzo[*a*]pyrene-7,8-dihydrodiol 9,10-epoxide adducts on adenine  $N^6$  at the second position of N-ras codon 61. *J. Biol. Chem.*, **270**, 4990–5000.
41. Frank, E.G., Sayer, J.M., Kroth, H., Ohashi, E., Ohmori, H., Jerina, D.M. and Woodgate, R. (2002) Translesion replication of benzo[*a*]pyrene and benzo[*c*]phenanthrene diol epoxide adducts of deoxyadenosine and deoxyguanosine by human DNA polymerase  $\iota$ . *Nucleic Acids Res.*, **30**, 5284–5292.
  42. Bigger, C.A., St. John, J., Yagi, H., Jerina, D.M. and Dipple, A. (1992) Mutagenic specificities of four stereoisomeric benzo[*c*]phenanthrene dihydrodiol epoxides. *Proc. Natl Acad. Sci. USA*, **89**, 368–372.
  43. Wolffe, W.T., Johnson, R.E., Minko, I.G., Lloyd, R.S., Prakash, S. and Prakash, L. (2006) Replication past a *trans*-4-hydroxynonenal minor-groove adduct by the sequential action of human DNA polymerases  $\iota$  and  $\kappa$ . *Mol. Cell Biol.*, **26**, 381–386.
  44. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
  45. Fiser, A., Do, R.K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.*, **9**, 1753–1773.
  46. Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461–491.
  47. Fiser, A. and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, **19**, 2500–2501.
  48. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R. and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
  49. Potapova, O., Chan, C., DeLucia, A.M., Helquist, S.A., Kool, E.T., Grindley, N.D. and Joyce, C.M. (2006) DNA polymerase catalysis in the absence of Watson–Crick hydrogen bonds: analysis by single-turnover kinetics. *Biochemistry*, **45**, 890–898.
  50. Mizukami, S., Kim, T.W., Helquist, S.A. and Kool, E.T. (2006) Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry*, **45**, 2772–2778.
  51. 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. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
  52. Cheatham, T.E. III, Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
  53. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T. *et al.* (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **24**, 1999–2012.
  54. Wang, J.M., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, **21**, 1049–1074.
  55. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
  56. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E. 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.*, **92**, 3817–3829.
  57. Williams, D.E. (1991) Net atomic charge and multipole models for the *ab initio* molecular electric potential. *Rev. Comp. Chem.*, **2**, 219–271.
  58. Reynolds, C.A., Essex, J.W. and Richards, W.G. (1992) Atomic charges for variable molecular conformations. *J. Am. Chem. Soc.*, **114**, 9075–9079.
  59. Hehre, W.J., Ditchfie, R. and Pople, J.A. (1972) Self-consistent molecular orbital methods. 12. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, **56**, 2257–2261.
  60. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A. Jr, Vreven, T., Kudin, K.N., Burant, J.C. *et al.* (2004) Gaussian, Inc., Wallingford, CT.
  61. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J. Phys. Chem.*, **97**, 10269–10280.
  62. Cieplak, P., Cornell, W.D., Bayly, C. and Kollman, P.A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers—charge derivation for DNA, RNA, and proteins. *J. Chem. Phys.*, **16**, 1357–1377.
  63. Perlow, R.A. and Broyde, S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: structural insights into nucleotide incorporation preferences. *J. Mol. Biol.*, **322**, 291–309.
  64. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
  65. Case, D.A., Darden, T.A., Cheatham, T.E. 3rd, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M. *et al.* (2006) University of California, San Francisco.
  66. Mezei, M. (1997) Optimal position of the solute for simulations. *J. Comp. Chem.*, **18**, 812–815.
  67. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
  68. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald—an  $n \log(n)$  method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
  69. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
  70. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical-integration of Cartesian equations of motion of a system with constraints—molecular dynamics of *N*-alkanes. *J. Comput. Phys.*, **23**, 327–341.
  71. Izaguirre, J.A., Catarello, D.P., Wozniak, J.M. and Skeel, R.D. (2001) Langevin stabilization of molecular dynamics. *J. Chem. Phys.*, **114**, 2090–2098.
  72. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
  73. Harvey, S.C., Tan, R.K.Z. and Cheatham, T.E. 3rd. (1998) The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.*, **19**, 726–740.
  74. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E. III, Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, A.L., Cheng, J.J. *et al.* (1999) *AMBER 6*. University of California, San Francisco, CA.
  75. Heyer, L.J., Kruglyak, S. and Yooseph, S. (1999) Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.*, **9**, 1106–1115.
  76. Bonneau, J.T.R., Ruczinski, I., Chivian, D., Rohl, C., Strauss, E.M.C. and Baker, D. (2001) Rosetta in CASP4: progress in *ab initio* protein structure prediction. *Proteins Struct. Funct. Genet.*, **45**, 119–126.
  77. Shao, J., Tanner, S.W., Thompson, N. and Cheatham, T.E. 3rd (2007) Clustering molecular dynamics trajectories: I. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.*, **3**, 2312–2334.
  78. Simmerling, C., Elber, R. and Zhang, J. (1995) MOIL-View—A Program for Visualization of Structure and Dynamics of Biomolecules and STO—A Program for Computing Stochastic Paths. In Pullman, A. *et al.* (eds), *Modeling of Biomolecular Structure and Mechanisms*. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 241–265.
  79. Wang, L., Yu, X., Hu, P., Broyde, S. and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for *Sulfolobus solfataricus* DNA polymerase IV. *J. Am. Chem. Soc.*, **129**, 4731–4737.
  80. Batra, V.K., Beard, W.A., Shock, D.D., Krahn, J.M., Pedersen, L.C. and Wilson, S.H. (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure*, **14**, 757–766.

81. Frank, E.G. and Woodgate, R. (2007) Increased catalytic activity and altered fidelity of human DNA polymerase  $\epsilon$  in the presence of manganese. *J. Biol. Chem.*, **282**, 24689–24696.
82. Boulard, Y., Cognet, J.A.H., Gabarro-Arpa, J., Le Bret, M., Carbonnaux, C. and Fazakerley, G.V. (1995) Solution structure of an oncogenic DNA duplex, the *K-ras* gene and the sequence containing a central C:A or A:G mismatch as a function of pH: nuclear magnetic resonance and molecular dynamics studies. *J. Mol. Biol.*, **246**, 194–208.
83. Lee, G.H. and Matsushita, H. (2005) Genetic linkage between Pol  $\epsilon$  deficiency and increased susceptibility to lung tumors in mice. *Cancer Sci.*, **96**, 256–259.
84. Ohkumo, T., Kondo, Y., Yokoi, M., Tsukamoto, T., Yamada, A., Sugimoto, T., Kanao, R., Higashi, Y., Kondoh, H., Tatematsu, M. *et al.* (2006) UV-B radiation induces epithelial tumors in mice lacking DNA polymerase  $\eta$  and mesenchymal tumors in mice deficient for DNA polymerase  $\epsilon$ . *Mol. Cell Biol.*, **26**, 7696–7706.
85. Dumstorf, C.A., Clark, A.B., Lin, Q., Kissling, G.E., Yuan, T., Kucherlapati, R., McGregor, W.G. and Kunkel, T.A. (2006) Participation of mouse DNA polymerase  $\epsilon$  in strand-biased mutagenic bypass of UV photoproducts and suppression of skin cancer. *Proc. Natl Acad. Sci. USA*, **103**, 18083–18088.
86. Wang, Y., Woodgate, R., McManus, T.P., Mead, S., McCormick, J.J. and Maher, V.M. (2007) Evidence that in xeroderma pigmentosum variant cells, which lack DNA polymerase  $\eta$ , DNA polymerase  $\epsilon$  causes the very high frequency and unique spectrum of UV-induced mutations. *Cancer Res.*, **67**, 3018–3026.
87. Ziv, O., Geacintov, N.E., Nakajima, S., Yasui, A. and Livneh, Z. (2009) DNA polymerase  $\zeta$  cooperates with polymerases  $\kappa$  and  $\epsilon$  in translesion DNA synthesis across pyrimidine photodimers in cells from XPV patients. *Proc. Natl Acad. Sci. USA*, **106**, 11552–11557.
88. Xie, X.-M., Geacintov, N.E. and Broyde, S. (1999) Stereochemical origin of opposite orientations in DNA adducts derived from enantiomeric *anti*-benzo[*a*]pyrene diol epoxides with different tumorigenic potentials. *Biochemistry*, **38**, 2956–2968.