

Article

# Computational Evaluation of Nucleotide Insertion Opposite Expanded and Widened DNA by the Translesion Synthesis Polymerase Dpo4

Laura Albrecht <sup>†</sup>, Katie A. Wilson <sup>†</sup> and Stacey D. Wetmore <sup>\*</sup>

Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge Alberta, AB T1K 3M4, Canada; laura.albrecht@uleth.ca (L.A.); ka.wilson@uleth.ca (K.A.W.)

<sup>\*</sup> Correspondence: stacey.wetmore@uleth.ca; Tel.: +1-403-329-2323; Fax: +1-403-329-2057

<sup>†</sup> These authors contributed equally to this work.

Academic Editors: James W. Gauld and Leif A. Eriksson

Received: 16 May 2016; Accepted: 14 June 2016; Published: 23 June 2016

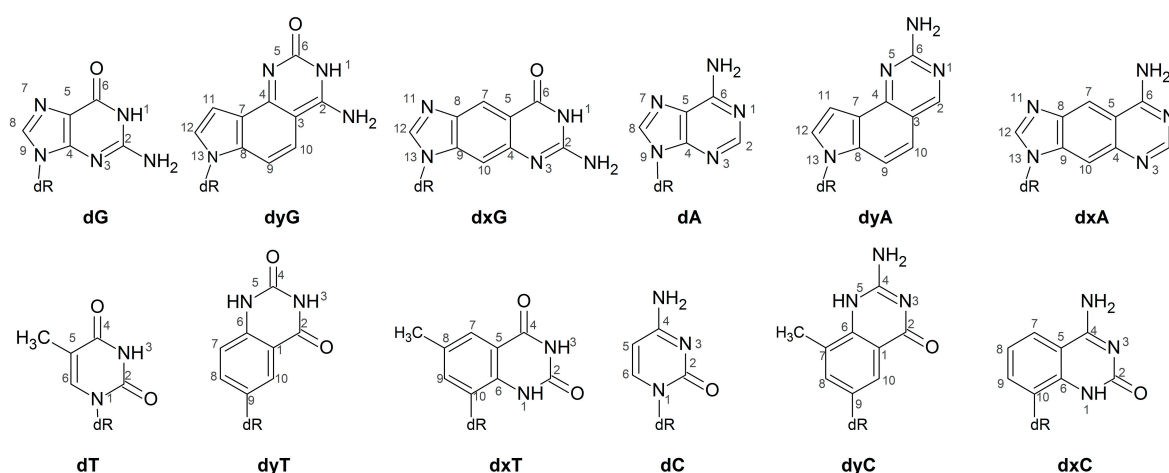
**Abstract:** Expanded (x) and widened (y) deoxyribose nucleic acids (DNA) have an extra benzene ring incorporated either horizontally (xDNA) or vertically (yDNA) between a natural pyrimidine base and the deoxyribose, or between the 5- and 6-membered rings of a natural purine. Far-reaching applications for (x,y)DNA include nucleic acid probes and extending the natural genetic code. Since modified nucleobases must encode information that can be passed to the next generation in order to be a useful extension of the genetic code, the ability of translesion (bypass) polymerases to replicate modified bases is an active area of research. The common model bypass polymerase DNA polymerase IV (Dpo4) has been previously shown to successfully replicate and extend past a single modified nucleobase on a template DNA strand. In the current study, molecular dynamics (MD) simulations are used to evaluate the accommodation of expanded/widened nucleobases in the Dpo4 active site, providing the first structural information on the replication of (x,y)DNA. Our results indicate that the Dpo4 catalytic (palm) domain is not significantly impacted by the (x,y)DNA bases. Instead, the template strand is displaced to accommodate the increased C1'–C1' base-pair distance. The structural insights unveiled in the present work not only increase our fundamental understanding of Dpo4 replication, but also reveal the process by which Dpo4 replicates (x,y)DNA, and thereby will contribute to the optimization of high fidelity and efficient polymerases for the replication of modified nucleobases.

**Keywords:** expanded DNA; xDNA; widened DNA; yDNA; DNA replication; translesion synthesis; bypass polymerase; Dpo4; molecular dynamics

## 1. Introduction

At the cellular level, living organisms store and process genetic information encoded by a set of four DNA nucleobases (A, C, G, and T), which combine according to Watson-Crick hydrogen-bonding rules. There have been many attempts to expand the unique function and applications of DNA [1,2]. Approaches considered to date include editing the DNA backbone (e.g., polyamide nucleic acids (PNA), locked nucleic acids (LNA), and xeno-nucleic acids (XNA) [3–5]), modifying the nucleobase functional groups (e.g., difluorotoluene, isoguanine, and 2,6-diaminopurine [6–9]), or changing the underlying nucleobase composition (e.g., propynyl isocarbostyryl and fleximers [10–12]). In addition to exploring fundamental questions about nature's building blocks, modified DNA has many practical applications, including DNA nanomaterials [13,14], therapeutic approaches such as antiviral drugs or the delivery of genetic material to cells [15–19], nucleic acid computing [20–23], and tools to probe the mechanisms of biological processes [24].

One class of modified DNA that shows promising applications both as an extension of the natural genetic code and as a nucleic acid probe is expanded (x) and widened (y) DNA. These modifications have a single benzene spacer incorporated horizontally (xDNA) or vertically (yDNA) either between a natural pyrimidine base and deoxyribose, or between the 5- and 6-membered rings of a natural purine (Figure 1). Although base extension increases the base-pair C1'–C1' distance, the canonical Watson-Crick hydrogen-bonding faces are maintained in (x,y)DNA nucleobases, which allows for complementary base pairing. The first benzene-ring expanded nucleobases were RNA analogues proposed by Leonard and coworkers [25–27]. The Kool group has since synthesized a series of expanded and widened DNA nucleobases [28–35]. Theoretical and experimental studies show that helices composed of (x,y)DNA:DNA base pairs are overall more stable than natural B-DNA [36,37], at least in part due to enhanced stacking [38], as well as exhibit increased charge-transfer [39–41] and fluorescent [36,42–45] properties. Compared to B-DNA, mixed xDNA–DNA helices have increased major and minor groove widths, and a reduced helical twist, while mixed yDNA–DNA helices have increased base-pair inclination, reduced twist, and smaller major and minor groove widths [34–36,41,42,46–49].



**Figure 1.** Chemical structures of the DNA, xDNA and yDNA bases considered in the current study.

In addition to the capacity to store information through complementary pairing in stable helices, it must be possible to replicate (x,y)DNA strands in order for these modifications to be useful extensions of the natural genetic code. In biological systems, DNA is copied by either standard replication or translesion (bypass) synthesis [50]. Standard replication uses replicative DNA polymerases that have high fidelity, processivity and efficiency for canonical nucleotides, but are generally unable to replicate non-standard (e.g., damaged or otherwise modified) bases [51,52]. On the other hand, bypass polymerases exploit a highly flexible active site to replicate past many non-standard bases, albeit with lower fidelity, processivity and efficiency compared to replication of canonical DNA by standard polymerases [50]. With the goal to improve the efficiency of modified DNA replication, several research groups have focused on engineering novel polymerase variants (by modifying individual residues and/or domains) that exploit the flexibility of bypass polymerases to replicate many non-standard bases with a range of biotechnological applications [53–55].

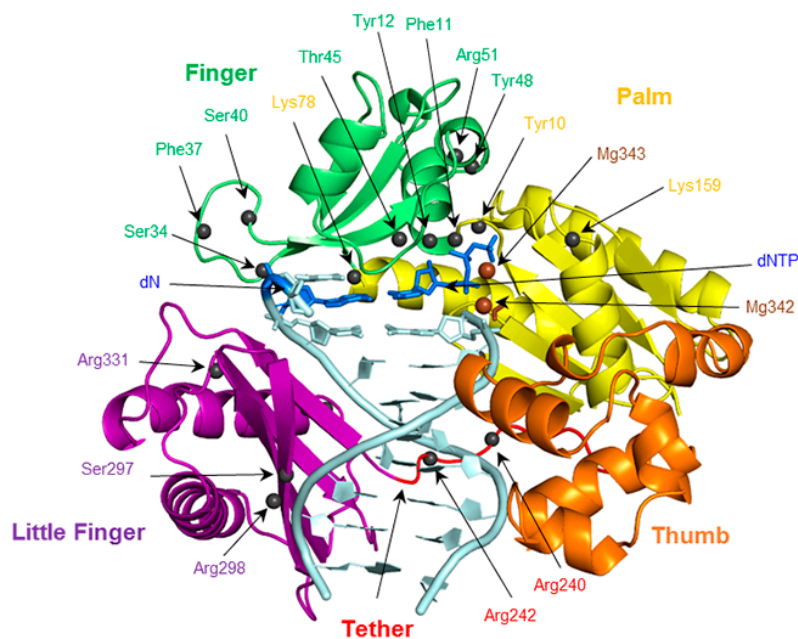
Since the design of an (x,y)DNA replicase is anticipated to be extremely challenging [33], experimental studies have initially investigated the replication of (x,y)DNA by existing polymerases. This approach was at least in part fuelled by the successful replication of DNA containing expanded and widened nucleobases in *Escherichia coli* (*E. coli*) [56–58]. To gain more detailed information, kinetic studies considered the replication of a single x- or y-nucleotide (dxN or dyN) in an otherwise canonical DNA duplex by common model replicative (Klenow fragment(*exo*-), Kf(*exo*-)) and bypass (Dpo4) polymerases [30,56,59]. These studies further confirmed successful insertion of a

complementary natural deoxyribose nucleoside triphosphate (dNTP) opposite  $d(x,y)N$  [30,59], but at a rate 2–3 orders of magnitude slower than complementary dNTP insertion opposite a canonical base for both polymerases [60]. Further study revealed that *Kf(exo-)* is unable to successfully extend past  $(x,y)DNA$  [30,59], whereas Dpo4 can extend past a single  $d(x,y)N$  paired with a canonical nucleobase [30,59], as well as replicate four consecutive  $xDNA$  bases [59]. Thus, replication of  $(x,y)DNA$  by Dpo4 appears to be the most promising avenue for future development considered in the literature so far. This is perhaps not surprising since Dpo4 has been shown to successfully bypass many damaged or synthetic bases, most having a large aromatic group added to a canonical nucleobase [50,61–74]. One unstudied exception is *anti*-purine:*anti*-purine mismatches, which have been speculated to impede Dpo4 nucleotide insertion by disrupting the alignment of the catalytic center because of increased  $C1'-C1'$  distances [64]. Unfortunately, there are no studies to date that provide structural details of the influence of base size expansion on the Dpo4 active site to better appreciate how base pairs with increased  $C1'-C1'$  distances are accommodated.

Since previous literature has demonstrated that computational studies provide critical molecular-level insight into the replication of non-standard nucleobases by Dpo4 [63,66,70,71,75,76], we have used molecular dynamics (MD) simulations to study Dpo4 replication of  $dxN$  or  $dyN$  nucleotides (Figure 1). Specifically, to align with previous experimental studies, we model a Dpo4 ternary complex for the insertion of a complementary canonical dNTP opposite an expanded ( $dxA$ ,  $dxC$ ,  $dxG$ ,  $dxT$ ), widened ( $dyA$ ,  $dyC$ ,  $dyG$ ,  $dyT$ ) or natural ( $dG$ ,  $dT$ ) base in an otherwise canonical DNA helix. Our simulations provide the first structural information about how the full set of expanded and widened DNA bases are accommodated in the Dpo4 active site. This data is crucial in order to better understand the process by which Dpo4 bypasses  $xDNA$  and  $yDNA$ , and thereby fully optimize a system that can replicate expanded and widened DNA with high efficiency and fidelity. Furthermore, our calculations provide the first information about structural deviations to the polymerase and DNA in the ternary (insertion) complex resulting from replication of modified bases that form pairs with increased  $C1'-C1'$  distances. When combined with previous studies on the replication of size-increased damaged DNA nucleobases that do not form complementary Watson-Crick hydrogen bonds, our work highlights the important role of hydrogen bonding for the replication of size-extended nucleobases. Thus, the present study significantly contributes to our growing fundamental understanding of the function of Dpo4, as well as other bypass polymerases.

## 2. Results

To provide the currently missing structural details of the replication of expanded and widened DNA bases, the present work systematically evaluates the conformation of the Dpo4 active site when bound to DNA containing a single  $(x,y)DNA$  base. Structural features to be considered include enzyme–dNTP interactions, the reaction parameters,  $d(x,y)N:dNTP$  hydrogen bonding, DNA structural (base-pair, base-step and helical) parameters associated with  $d(x,y)N:dNTP$  and the previously replicated base pairs, and enzyme–dN interactions. Global changes in the Dpo4 ternary complex, including characterizing any potentially significant rearrangements in each domain (palm, finger, thumb and little finger; Figure 2), as well as the tether region, are also considered. The geometrical features of the Dpo4 ternary complex for the replication of expanded or widened bases are compared with the insertion complex for natural DNA replication. Previous research indicates that canonical purine and pyrimidine dNTPs may themselves be accommodated differently in the Dpo4 active site. For example, crystal structures of the Dpo4 insertion complex indicate that stabilizing enzyme–DNA interactions vary for dCTP insertion opposite dG compared to dATP insertion opposite dT [77,78]. To account for these potential differences, the results for the replication of a modified purine or pyrimidine are compared to the corresponding Dpo4 ternary complex for the replication of either a natural purine (dG) or pyrimidine (dT), respectively. Due to the large amount of data generated, and similarities across the systems considered, the results section primarily focuses on the most significant differences between the systems.

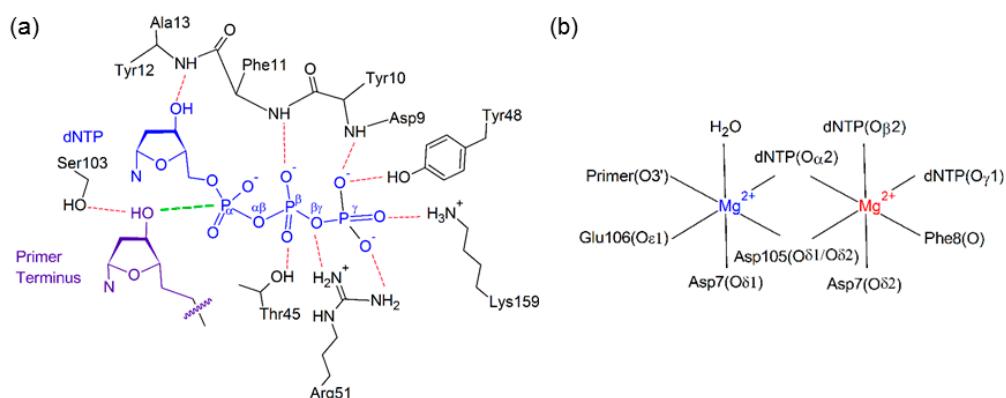


**Figure 2.** The finger, palm, thumb, and little finger domains of Dpo4, as well as the tether region. The positions of potentially important residues in each domain are indicated with spheres.

### 2.1. Catalytic Palm Domain Does Not Significantly Rearrange for the Replication of (x,y)DNA Bases, with the Exception of Modified Cytosine

Stabilization of the dNTP and primer strand terminus plays an important role in aligning the active center of Dpo4, which sits inside the palm domain (Figure 2). Indeed, previous computational and experimental studies have identified several interactions that position the dNTP in the Dpo4 active site (Figure 3). These include hydrogen bonds with two residues in the palm domain (Tyr10 and Lys159) and five residues in the finger domain (Thr45, Phe11, Arg51, Tyr12, and Tyr48), as well as electrostatic interactions with two  $Mg^{2+}$  ions [63,70,76,79–81]. Although most enzyme–dNTP interactions observed for canonical DNA replication (Figure 3a) are conserved upon base modification, there are some changes in these enzyme–dNTP interactions for the replication of (x,y)DNA nucleobases compared to the canonical counterparts, particularly for the replication of benzene-extended cytosines (Table 1 and Tables S1–S3). For example, for the replication of canonical dT, both dATP(O3′H) and Thr45(O $\gamma$ H) form a hydrogen bond with dATP(O $\beta$ ). However, for the replication of the modified cytosines, the corresponding hydrogen-bonding interaction with Thr45 is completely lost (Table 1), leading to a destabilization of about 8 kcal·mol<sup>−1</sup> according to MM/GBSA binding energies (Table S1). Furthermore, the hydrogen bonding with Arg51 is lost due to a severely narrowed dNTP(O $\beta$  $\gamma$ )···Arg51(N $\eta$ 1H) hydrogen-bond angle (135° for dyC and 116° for dxC), which results in an ~10 kcal·mol<sup>−1</sup> reduction in stability (Table S1). Finally, the hydrogen bonding between Tyr10(NH) and dNTP(O $\gamma$ ) is reduced (occupancy decreases by ~40%; Table 1), although the associated destabilization is relatively small (~3 kcal·mol<sup>−1</sup>). In addition to these deviations for the modified cytosine insertion complexes, some changes occur when the replication of the modified purines are considered. For example, although there is a stabilizing interaction between the O3′ hydroxy group of dCTP and the aromatic ring of Tyr12 for the replication of dG, the O3′ hydroxy in dNTP points away from Tyr12 and towards Thr45 for the replication of dxA and dxG (Figure S1). In the dxG, dyG, dyA and dxT insertion complexes, the Lys159(N $\zeta$ H)···dNTP(O $\gamma$ ) hydrogen-bond occupancy is reduced by ~20%, but only by 10% for dyT. Finally, the aromatic ring of Tyr48 rotates by ~45° to potentially form a stacking interaction with Arg51 when dxG or dyA are bound in the Dpo4 active site, which is not observed in the natural systems (Figure S2).

The Dpo4 active center has two hexa-coordinated  $Mg^{2+}$  ions (Figure 3b), one (catalytic  $Mg^{2+}$ ) that stabilizes the 3'-terminal hydroxy group of the DNA primer strand and a second (binding  $Mg^{2+}$ ) that stabilizes the triphosphate moiety [76,77]. Coordination of these important ions is unaffected by the modified bases, with all six coordinating bonds having greater than 99% occupancy for a distance of  $<2.5 \text{ \AA}$  (Table S4). We also observe a hydrogen-bonding interaction with the primer strand terminus that has not been previously reported; specifically, Ser103 hydrogen bonds with the 3'-hydroxy group of the primer strand terminus for  $\sim 90\%$  of the simulation trajectory for the natural systems (Table S3). This occupancy decreases by 20% for dyA, and by 10% for dxC, dyC, and dxA.



**Figure 3.** (a) Hydrogen-bonding interactions with the dNTP (blue) and primer terminus (purple), with the reaction distance shown in green; (b) Active site coordination of the catalytic (blue) and binding (red)  $Mg^{2+}$  ions.

**Table 1.** Percent occupancies for hydrogen-bonding interactions between the dNTP and active site residues <sup>a</sup>.

Active Site Base Pair	dNTP(O $\beta$ $\gamma$ ) Arg51(N $\eta$ 1H)	dNTP(O $\gamma$ ) Arg51(N $\eta$ 2H)	dNTP(O $\gamma$ ) Lys159(N $\zeta$ H)	dNTP(O $\gamma$ ) Tyr10(NH)	dNTP(O $\beta$ ) Thr45(O $\gamma$ H)	dNTP(O $\beta$ ) dNTP(O3'H)
dG:dCTP	100%	97%	90%	98%	99%	8%
dT:dATP	98%	99%	88%	99%	96%	99%
dxA:dTTP	98%	100%	93%	96%	98%	11%
dyA:dTTP	93%	100%	76%	98%	100%	NA <sup>b</sup>
dxG:dCTP	100%	90%	68%	98%	97%	6%
dyG:dCTP	99%	90%	71%	98%	95%	NA <sup>b</sup>
dxT:dATP	99%	97%	61%	99%	96%	100%
dyT:dATP	99%	98%	78%	99%	96%	100%
dxC:dGTP	NA <sup>b</sup>	100%	99%	44%	NA <sup>b</sup>	93%
dyC:dGTP	NA <sup>b</sup>	100%	99%	41%	NA <sup>b</sup>	93%

<sup>a</sup> Hydrogen-bonding occupancies are based on a distance cutoff of  $<3.4 \text{ \AA}$  and an angle cutoff of  $>120^\circ$ . Table S2 also contains average distances and angles of hydrogen-bonding interactions with the dNTP; <sup>b</sup> Not observed with  $>5\%$  occupancy within our hydrogen-bonding criteria.

## 2.2. Reaction Parameters for the Catalytic Step Are Unaffected for the Replication of All (x,y)DNA Bases

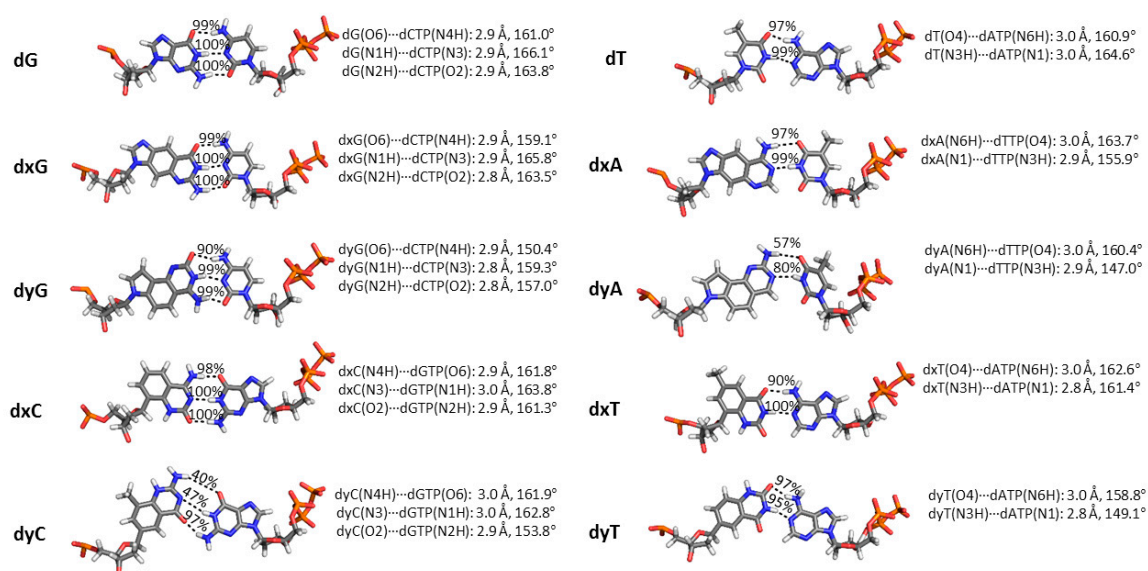
The catalytic insertion step involves water-mediated O3' attack on P $\alpha$ , formation of a pentavalent intermediate, and finally breaking of the P $\alpha$ -O $\alpha$  $\beta$  bond in the dNTP to free pyrophosphate from the active site [76,82]. Two important parameters that indicate correct alignment for the insertion reaction are the distance between O3' of the primer strand terminus and the  $\alpha$ -phosphate of the dNTP, and the angle between these atoms and O $\alpha$  $\beta$  of the dNTP. The results show that the dxN and dyN bases have little effect on the alignment of the Dpo4 catalytic center (Table 2). Specifically, the average O3'-P $\alpha$  distance is within 0.1  $\text{\AA}$  of the natural systems and the average O3'-P $\alpha$ -O $\alpha$  $\beta$  reaction angle is within  $11^\circ$  of the natural systems.

**Table 2.** Average and standard deviation of the reaction distance and angle during 40 ns MD simulations of the Dpo4 ternary complex for the replication of a DNA, xDNA or yDNA base.

Active Site Base Pair	d(O3'–P $\alpha$ )	$\angle$ (O3'–P $\alpha$ –O $\alpha\beta$ )
dG:dCTP	3.6 $\pm$ 0.1 Å	172.9 $\pm$ 3.7°
dT:dATP	3.5 $\pm$ 0.1 Å	174.0 $\pm$ 3.0°
dxA:dTTP	3.6 $\pm$ 0.1 Å	174.4 $\pm$ 3.0°
dyA:dTTP	3.6 $\pm$ 0.2 Å	174.2 $\pm$ 3.1°
dxG:dCTP	3.6 $\pm$ 0.1 Å	174.1 $\pm$ 3.2°
dyG:dCTP	3.6 $\pm$ 0.1 Å	174.2 $\pm$ 3.1°
dxT:dATP	3.5 $\pm$ 0.1 Å	174.1 $\pm$ 3.1°
dyT:dATP	3.5 $\pm$ 0.1 Å	173.7 $\pm$ 3.3°
dxC:dGTP	3.6 $\pm$ 0.2 Å	163.1 $\pm$ 4.5°
dyC:dGTP	3.6 $\pm$ 0.2 Å	162.6 $\pm$ 4.5°

### 2.3. *d(x,y)N:dNTP Hydrogen Bonding Does Not Significantly Deviate from Canonical Watson-Crick Interactions, with the Exception of dyA and dyC*

In light of the ongoing literature debate regarding the role of hydrogen bonding in dNTP insertion by translesion polymerases [83–86], it is essential to monitor interactions between the template dN and the incoming dNTP (Figure 4). In the Dpo4 active site, dG:dCTP and dT:dATP form canonical Watson–Crick hydrogen bonds with greater than 97% occupancy. Although the Watson–Crick hydrogen-bonding pattern is maintained for all *d(x,y)N:dNTP* pairs, the interactions significantly change for *dyA* and *dyC*. Specifically, the average hydrogen-bond angles in the *dyA:dTTP* pair decrease by up to  $\sim 20^\circ$  (hydrogen-bond occupancy decreases by up to 43%). For the replication of *dyC*, two hydrogen bonds are disrupted during the simulation (up to 60% decrease in occupancy). The next largest structural change occurs between the *dyG*(O6) acceptor and dCTP(N4H) donor, with the average hydrogen-bond angle narrowing by up to  $\sim 10^\circ$  (10% decrease in occupancy). For the *n*–1 dG:dC base pair, only dxT and dyC led to a small geometric deviation from the natural systems (6% decrease in the dC(N4H)···dG(O6) hydrogen-bonding occupancy; Table S5).



**Figure 4.** Geometry of the dN:dNTP hydrogen-bonding interaction during Dpo4 replication of a DNA, xDNA or yDNA base. Average distance and angle for each hydrogen-bonding interaction are given, as well as hydrogen-bonding occupancies, which are based on a distance cutoff of  $<3.4$  Å and an angle cutoff of  $>120^\circ$ .

#### 2.4. Base-Pair, Base-Step and Helical Parameters for (x,y)DNA Deviate from Canonical Values, with Widened Bases Generally Resulting in Greater Deviations than Expanded Bases

Expanded and widened DNA helices have been shown to adopt a native B-DNA geometry with a reduced twist, as well as increased (decreased) major and minor groove widths for xDNA (yDNA) [30]. Therefore, the geometry of the DNA helix (base-pair, base-step and helical parameters) [87,88] in the Dpo4 ternary complex for replication of an expanded or widened base was evaluated. The most significant deviation from the natural complexes in the base-pair parameters that describe the relative orientation of the d(x,y)N and dNTP bases occurs in the buckle parameter (Table 3). Specifically, dyA:dTTP, dyG:dCTP and dyT:dATP exhibit an increase in the average buckle up to  $\sim 30^\circ$ , while dxA:dTTP and dxG:dCTP have an average increase in buckle of  $\sim 15^\circ$ , relative to the natural systems (Figures S3 and S4). The average propeller rotation deviates by up to  $\sim 15^\circ$  from the natural systems for dyA:dTTP and dyC:dGTP. The average opening of the dN:dNTP base pair is unchanged with respect to the natural complexes, with the exception of dyC:dGTP (average increases by  $\sim 14^\circ$ ) and dyA:dTTP (which has a large standard deviation of  $12.2^\circ$ ). Due to the presence of the additional benzene ring in the modified bases, a large increase in stretch is observed for all complexes (Figure 5). The average stretch increases by 1.2–2.4 Å for the widened bases, but falls within a narrower range for the expanded bases (1.8–2.3 Å). Correlated to this increase in stretch, the average C1'–C1' distance increases by  $\sim 1$ –2 Å compared to the natural systems. Shear displacement increases for dyA, dyG, dyT and dxC, with an average increase compare to the natural systems of up to  $\sim 2.6$  Å. Finally, only the modified purines lead to a notable increase in average stagger (by up to 1.5 Å).

Among the base-step parameters that dictate the relative orientation between the dN:dNTP and n–1 dG:dC pair [87,88], only the tilt and twist are significantly affected by the presence of a modified base, and these parameters only deviate for the widened purine systems (Table 4). Specifically, the average tilt increases for dyA and dyG by up to  $\sim 10^\circ$ , and the average twist for dyA is  $10^\circ$  larger than for natural DNA replication. The inclination and tip helical parameters are overall unchanged for the d(x,y)N base pairs with the exception of a  $\sim 12^\circ$  increase in average tip for dyA:dTTP, and large standard deviations ( $\sim 15^\circ$ ) for the inclination and tip of dxC:dGTP. Although the major groove width remains consistent across all natural and modified bases (Table 4), dyA leads to significant changes in the average minor groove width (1.6 Å decrease) and axial bend (1.7 Å increase), with associated standard deviations also being larger than for any other system. The rest of the DNA helix shows no significant changes in base-pair, base-step or helical parameters (Figures S5 and S6).

**Table 3.** Averages and standard deviations of the base-pair parameters during 40 ns MD simulations of the Dpo4 ternary complex for the replication of a DNA, xDNA or yDNA base <sup>a,b</sup>.

Active Site Base Pair	Buckle	Propeller	Opening	Shear	Stretch	C1'–C1' Distance	Stagger
dG:dCTP	$-5.3 \pm 7.4$	$-1.2 \pm 7.1$	$0.8 \pm 2.9$	$0.2 \pm 0.3$	$0.0 \pm 0.1$	$10.8 \pm 0.1$	$-0.3 \pm 0.4$
dT:dATP	$1.5 \pm 7.8$	$-4.7 \pm 7.1$	$0.6 \pm 4.7$	$-0.1 \pm 0.3$	$0.1 \pm 0.1$	$10.8 \pm 0.2$	$-0.1 \pm 0.4$
dxA:dTTP	$-14.2 \pm 9.3$	$-9.5 \pm 7.1$	$5.5 \pm 5.0$	$0.4 \pm 0.3$	$2.2 \pm 0.2$	$12.6 \pm 0.2$	$-1.2 \pm 0.7$
dyA:dTTP	$-28.7 \pm 13.6$	$-19.8 \pm 9.5$	$3.7 \pm 12.2$	$-2.6 \pm 0.6$	$1.2 \pm 0.3$	$11.9 \pm 0.5$	$-1.9 \pm 0.9$
dxG:dCTP	$-18.2 \pm 9.6$	$-6.0 \pm 7.2$	$1.5 \pm 3.5$	$0.4 \pm 0.3$	$2.3 \pm 0.1$	$12.8 \pm 0.2$	$-1.3 \pm 0.6$
dyG:dCTP	$-30.6 \pm 10.2$	$-10.1 \pm 7.9$	$-3.3 \pm 3.5$	$-1.3 \pm 0.4$	$1.7 \pm 0.1$	$12.4 \pm 0.2$	$-1.7 \pm 0.7$
dxT:dATP	$-8.5 \pm 7.8$	$-7.2 \pm 9.3$	$1.3 \pm 5.9$	$0.4 \pm 0.3$	$1.8 \pm 0.2$	$12.2 \pm 0.2$	$-0.4 \pm 0.5$
dyT:dATP	$-24.2 \pm 13.1$	$-2.7 \pm 7.5$	$5.1 \pm 3.7$	$-1.2 \pm 0.3$	$1.9 \pm 0.2$	$12.2 \pm 0.3$	$-0.1 \pm 0.7$
dxC:dGTP	$3.3 \pm 6.8$	$-6.3 \pm 9.1$	$3.1 \pm 3.8$	$1.2 \pm 0.3$	$2.1 \pm 0.1$	$12.7 \pm 0.2$	$0.3 \pm 0.4$
dyC:dGTP	$0.1 \pm 9.3$	$-15.1 \pm 8.5$	$13.8 \pm 14.0$	$-0.3 \pm 0.9$	$2.4 \pm 0.4$	$12.4 \pm 0.5$	$-0.2 \pm 0.4$

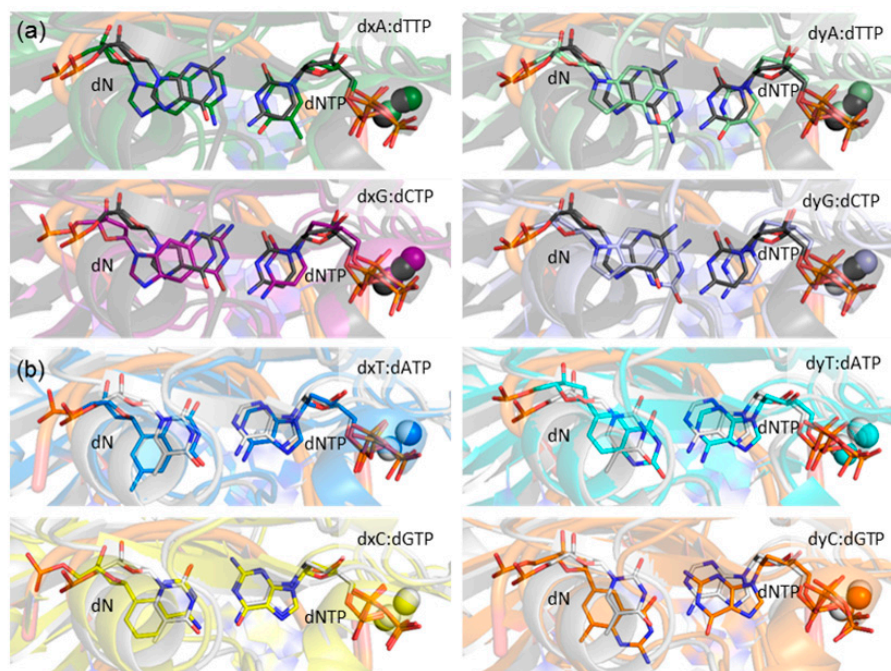
<sup>a</sup> See Figure S5 for the base-pair parameters for the rest of the helix; <sup>b</sup> Buckle, propeller and opening are in degrees, while shear, stretch, C1'–C1' distance and stagger are in Angstroms.

**Table 4.** Averages and standard deviations of the base-step and helical parameters during 40 ns MD simulations of the Dpo4 ternary complex for the replication of a DNA, xDNA or yDNA base <sup>a,b</sup>.

Active Site Base Pair	Tilt	Roll	Twist	Shift	Slide	Rise	Inclination	Tip	Major Groove Width	Minor Groove Width	Axial Bend
dG:dCTP	$-4.1 \pm 3.5$	$7.7 \pm 4.1$	$26.3 \pm 3.1$	$0.2 \pm 0.4$	$-0.8 \pm 0.3$	$3.2 \pm 0.2$	$18.2 \pm 4.5$	$0.1 \pm 4.4$	$14.0 \pm 1.4$	$8.3 \pm 1.8$	$2.2 \pm 1.0$
dT:dATP	$-0.7 \pm 4.0$	$5.0 \pm 4.7$	$29.8 \pm 3.1$	$0.4 \pm 0.5$	$-0.9 \pm 0.3$	$3.2 \pm 0.3$	$14.0 \pm 5.7$	$0.1 \pm 4.2$	$13.4 \pm 1.6$	$8.0 \pm 2.0$	$2.0 \pm 0.9$
dxA:dTTP	$-7.7 \pm 4.2$	$4.9 \pm 3.8$	$21.9 \pm 2.8$	$0.1 \pm 0.5$	$-1.2 \pm 0.3$	$3.1 \pm 0.2$	$17.8 \pm 5.7$	$4.6 \pm 4.5$	$13.2 \pm 0.9$	$7.8 \pm 1.7$	$2.2 \pm 1.1$
dyA:dTTP	$-17.3 \pm 7.5$	$4.8 \pm 4.6$	$36.1 \pm 4.0$	$0.0 \pm 0.7$	$0.0 \pm 0.8$	$2.9 \pm 0.3$	$23.3 \pm 6.5$	$12.4 \pm 5.5$	$13.5 \pm 1.0$	$6.4 \pm 3.7$	$3.7 \pm 2.0$
dxG:dCTP	$-8.0 \pm 4.3$	$6.2 \pm 4.1$	$22.1 \pm 3.0$	$-0.1 \pm 0.4$	$-1.2 \pm 0.3$	$3.1 \pm 0.2$	$20.5 \pm 5.3$	$3.9 \pm 4.6$	$13.3 \pm 1.0$	$7.6 \pm 2.7$	$3.2 \pm 1.5$
dyG:dCTP	$-12.4 \pm 5.2$	$4.6 \pm 3.7$	$28.2 \pm 3.0$	$-0.1 \pm 0.4$	$-0.6 \pm 0.4$	$3.0 \pm 0.3$	$20.2 \pm 4.8$	$8.6 \pm 4.6$	$13.2 \pm 1.1$	$8.5 \pm 1.9$	$2.1 \pm 1.1$
dxT:dATP	$-1.8 \pm 4.0$	$5.8 \pm 4.4$	$25.1 \pm 3.2$	$0.3 \pm 0.5$	$-1.4 \pm 0.3$	$3.1 \pm 0.2$	$15.0 \pm 5.4$	$-1.1 \pm 4.3$	$13.5 \pm 1.0$	$8.1 \pm 1.3$	$2.7 \pm 1.2$
dyT:dATP	$-4.4 \pm 4.8$	$1.4 \pm 5.2$	$29.9 \pm 3.0$	$-0.2 \pm 0.4$	$-0.9 \pm 0.4$	$3.3 \pm 0.3$	$12.0 \pm 5.3$	$4.1 \pm 5.0$	$13.3 \pm 1.2$	$8.2 \pm 1.4$	$2.5 \pm 1.2$
dxC:dGTP	$-5.8 \pm 3.3$	$5.7 \pm 4.1$	$19.8 \pm 3.4$	$0.2 \pm 0.4$	$-1.5 \pm 0.3$	$3.4 \pm 0.2$	$19.7 \pm 14.6$	$-1.9 \pm 14.7$	$13.1 \pm 0.9$	$7.9 \pm 1.1$	$2.6 \pm 1.1$
dyC:dGTP	$-6.8 \pm 3.8$	$10.0 \pm 5.0$	$26.5 \pm 3.5$	$-0.6 \pm 0.8$	$-0.9 \pm 0.4$	$3.3 \pm 0.2$	$22.0 \pm 7.0$	$0.9 \pm 4.6$	$13.6 \pm 1.2$	$8.4 \pm 2.0$	$1.9 \pm 1.0$

<sup>a</sup> See Figure S6 for the base-step parameters for the rest of the helix; <sup>b</sup> Tilt, roll, twist, inclination, tip and axial bend are in degrees, while shift, slide, rise and major/minor groove widths are in Angstroms.





**Figure 5.** Overlays of representative MD structures of the (x,y)DNA base pairs with representative MD structures of the (a) dG:dCTP (black) or (b) dT:dATP (white) pair in the Dpo4 ternary complexes.

### 2.5. Canonical dN:dNTP Base-Pair Interactions Are Maintained for (x,y)DNA Bases, with the Exception of dyC, dxC and dyA

In order to appreciate the energetic effects of the base-pair and base-step distortions, the average linear interaction energies were calculated for the dN:dNTP and  $n-1$  dC:dG pairs across the simulation trajectory (Table 5). For natural DNA replication, the dG:dCTP Watson-Crick hydrogen-bonding interaction has an electrostatic energy of  $-30.7 \pm 2.7$  kcal·mol<sup>-1</sup>, which is slightly higher than previously estimated using gas-phase calculations on an isolated G:C base pair ( $-25$  to  $-28$  kcal mol<sup>-1</sup> [89]). The dT:dATP interaction is  $-7.1 \pm 1.6$  kcal·mol<sup>-1</sup>, which is  $\sim 5$  kcal mol<sup>-1</sup> lower than previously estimated for an isolated A:T base pair ( $-12$  kcal·mol<sup>-1</sup> [89]). The hydrogen-bond interaction in the  $n-1$  dG:dC pair maintains the strength of a canonical pair, ranging between  $-27.6$  and  $-28.8$  kcal·mol<sup>-1</sup> regardless of the base being replicated. Although the majority of the (x,y)DNA pairs maintain the stability of their canonical counterparts, dyC:dGTP is significantly less stable, with a hydrogen-bond energy of only  $-4.5 \pm 2.9$  kcal·mol<sup>-1</sup>, as expected due to changes in the base pair geometry. The only other significant differences are the reduced hydrogen-bonding energy for dxC:dGTP ( $-23.4 \pm 2.6$  kcal·mol<sup>-1</sup>) and increased standard deviation for dyA:dTTP ( $-5.8 \pm 3.5$  kcal·mol<sup>-1</sup>).

The stacking interactions between the d(x,y)N:dNTP and  $n-1$  dC:dG pairs reported herein (Table 5) are within 1–2 kcal·mol<sup>-1</sup> of the empirical (AMBER) van der Waals energies of isolated stacked base pairs [90]. Indeed, the stacking energies for the expanded bases are not significantly greater than for the natural bases, and the stacking energies for the expanded and widened bases are consistently within 1 kcal·mol<sup>-1</sup>. The (x,y)purine stacking energies with the  $n-1$  dC are on average 2 kcal·mol<sup>-1</sup> more stabilizing than the analogous (x,y)pyrimidine interactions. Interestingly, dxT and dyC have stabilizing stacking interactions with the  $n+1$  dC, largely attributed to an interaction between the methyl group in dxT or dyC and the  $n+1$  dC (Figure S7). This methyl interaction anchors the  $n+1$  dC above the modified base, which inhibits the  $n+1$  dC from rotating away from the reaction center as previously reported in crystal structures of natural and damaged DNA [91], and observed for all other modified bases in the present work.

**Table 5.** AMBER linear interaction energies (kcal·mol<sup>-1</sup>) for the dN:dNTP base pair during 40 ns MD simulations of the Dpo4 ternary complex for the replication of a DNA, xDNA or yDNA base <sup>a</sup>.

Active Site Base Pair	dN:dNTP Hydrogen Bonding	dC:dG n-1 Hydrogen Bonding	dNTP-dG n-1 Stacking	dN-dC n-1 Stacking	dN-dC n+1 Stacking	dN:dNTP-dC:dG n-1 Stacking
dG:dCTP	-30.7 ± 2.7	-29.3 ± 2.5	-6.1 ± 0.6	-6.6 ± 0.6	-0.3 ± 0.2	-15.4 ± 1.0
dT:dATP	-7.0 ± 1.6	-28.8 ± 2.7	-7.1 ± 0.6	-5.1 ± 0.7	-0.4 ± 0.2	-15.1 ± 0.9
dxA:dTTP	-9.0 ± 2.0	-27.9 ± 3.0	-6.6 ± 0.6	-7.3 ± 0.7	-0.4 ± 0.3	-16.8 ± 1.0
dyA:dTTP	-5.8 ± 3.5	-27.6 ± 3.0	-5.5 ± 0.9	-6.9 ± 0.7	-0.4 ± 0.4	-15.6 ± 1.3
dxG:dCTP	-26.9 ± 2.7	-28.3 ± 2.8	-6.0 ± 0.6	-7.6 ± 0.7	-1.5 ± 1.1	-16.8 ± 1.0
dyG:dCTP	-31.4 ± 3.8	-28.0 ± 2.7	-5.7 ± 0.6	-6.9 ± 0.7	-0.5 ± 0.4	-15.7 ± 1.0
dxT:dATP	-9.4 ± 1.6	-27.6 ± 3.2	-6.9 ± 0.7	-4.6 ± 0.7	-1.9 ± 1.4	-14.0 ± 1.1
dyT:dATP	-9.3 ± 1.7	-27.6 ± 3.0	-6.9 ± 0.8	-4.1 ± 0.8	-0.6 ± 0.7	-13.4 ± 1.2
dxC:dGTP	-23.4 ± 2.6	-27.8 ± 3.0	-7.2 ± 0.7	-5.5 ± 0.7	0.0 ± 0.1	-15.6 ± 1.1
dyC:dGTP	-4.5 ± 2.9	-27.9 ± 2.8	-7.6 ± 0.8	-5.4 ± 0.9	-3.3 ± 1.7	-14.4 ± 1.3

<sup>a</sup> The hydrogen-bonding energies are reported as the electrostatic component, while the stacking interactions are reported as the van der Waals energy component.

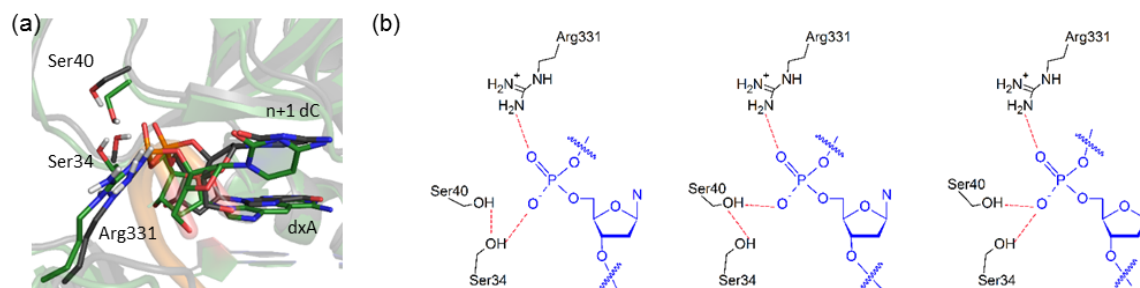
## 2.6. Residues Interacting with the Template Backbone Move to Accommodate the (x,y)DNA Bases

The largest structural deviations in the Dpo4 ternary complexes for (x,y)DNA versus canonical DNA replication occur near the expanded or widened base. Specifically, the extended base size causes the backbone of the template strand to shift outward (i.e., in the direction of an increased stretch parameter, see Figure 5). This shift is on average ~2.5 Å for the expanded and widened pyrimidines, except for dxC where the backbone shifts on average 3.0 Å (Figure 5 and Table 6). The expanded and widened purines lead to smaller average movement of 1.2–1.8 Å. Regardless, this motion directly impacts stabilizing interactions between the protein and the backbone of the template strand, particularly contacts involving Lys78, Arg331, Ser34, and Ser40. Specifically, in the natural complexes, there is a stabilizing interaction (3–7 kcal·mol<sup>-1</sup>; Table S6) between Lys78 (palm domain) and the template backbone at the n-2 position that is present for up to 53% of the simulations (Table S7). This interaction is lost in all modified systems. Arg331 forms a hydrogen-bonding interaction with the backbone of the n+1 dC for natural DNA replication (Figure 6a and Table S8). Although this interaction is still present for dxN or dyN replication, the increase in base-pair size shifts Arg331, causing the α-carbon of the protein backbone to move outwards by up to an average distance of 1.4 Å for the modified systems (Figure S8). Finally, in the natural systems, the hydroxy groups of Ser34 and Ser40 form a cooperative hydrogen-bond chain with the dN phosphate moiety (i.e., Ser40(OγH)···Ser34(OγH)···dN(OP); Figure 6b, left). Although this hydrogen-bonding pattern is maintained for dxC replication, the hydrogen-bonding chain switches direction in the dyG complex (Figure 6b, middle). In contrast, Ser40 is directed away from the helix backbone and forms a Phe37(NH)···Ser40(Oγ) hydrogen bond for the majority of the dxG simulation. For all other (x,y)DNA bases, both Ser34 and Ser40 form direct hydrogen bonds with the same d(x,y)N(OP) atom (Figure 6b, right).

**Table 6.** RMSD (Å) in the backbone position in the Dpo4 ternary complex for the replication of a (x,y)DNA base compared to a natural base across the 40 ns MD simulation trajectories <sup>a</sup>.

Active Site Base Pair	dN	Template n-1	dNTP	Primer n-1	Ser34	Ser40	Gly58	Met76	Arg331
dxA:dTTP	1.7 ± 0.5	1.0 ± 0.3	0.5 ± 0.1	0.7 ± 0.5	1.1 ± 0.4	1.1 ± 0.4	0.8 ± 0.4	0.7 ± 0.3	1.4 ± 0.4
dyA:dTTP	1.2 ± 0.4	0.7 ± 0.3	0.4 ± 0.1	0.8 ± 0.5	1.1 ± 0.4	1.0 ± 0.4	0.9 ± 0.4	0.7 ± 0.3	0.8 ± 0.4
dxG:dCTP	1.8 ± 0.4	1.0 ± 0.3	0.5 ± 0.1	0.7 ± 0.4	1.1 ± 0.4	0.9 ± 0.5	0.9 ± 0.4	0.7 ± 0.3	1.1 ± 0.4
dyG:dCTP	1.4 ± 0.3	0.9 ± 0.3	0.5 ± 0.2	0.8 ± 0.3	1.1 ± 0.4	1.1 ± 0.4	0.8 ± 0.4	0.7 ± 0.3	1.2 ± 0.4
dxT:dATP	3.0 ± 0.8	1.4 ± 0.4	0.4 ± 0.2	0.8 ± 0.4	0.9 ± 0.4	1.0 ± 0.5	1.0 ± 0.4	0.7 ± 0.3	1.3 ± 0.4
dyT:dATP	2.6 ± 1.0	1.0 ± 0.4	0.5 ± 0.2	0.8 ± 0.4	1.1 ± 0.5	1.2 ± 0.6	1.0 ± 0.4	0.8 ± 0.4	1.3 ± 0.5
dxC:dGTP	2.6 ± 0.9	0.8 ± 0.4	0.4 ± 0.2	0.8 ± 0.4	1.3 ± 0.4	1.1 ± 0.4	1.3 ± 0.5	0.9 ± 0.4	1.1 ± 0.4
dyC:dGTP	2.5 ± 0.8	0.8 ± 0.3	0.4 ± 0.1	0.8 ± 0.4	1.1 ± 0.4	1.2 ± 0.5	1.2 ± 0.5	0.7 ± 0.3	1.2 ± 0.4

<sup>a</sup> Backbone position was monitored according to P for a DNA component or Cα for a protein component.



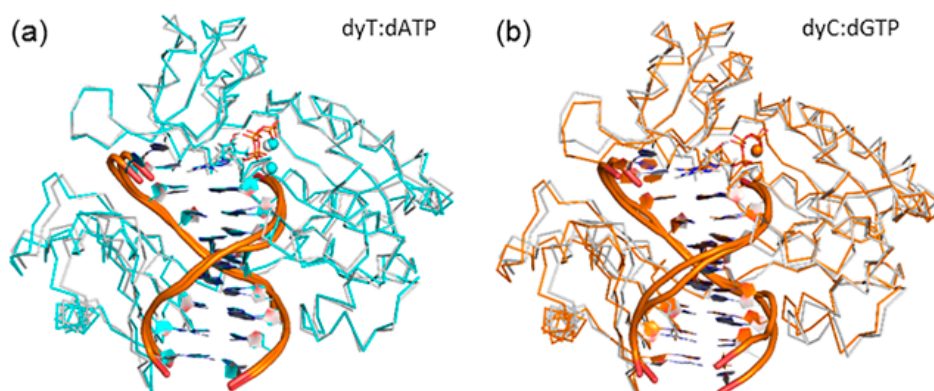
**Figure 6.** Alterations in the Dpo4 conformation to accommodate an (x,y)DNA base in the template strand. (a) Overlay of the representative MD structures of the Dpo4 ternary complex for the replication of dxA (green) and dG (black). Overlays for all systems are provided in the Supplementary Material (Figure S8); (b) Possible hydrogen-bonding arrangements within the Ser40...Ser34...d(x,y)N chain.

### 2.7. Residues in the Ceiling Region of the Finger Domain Rearrange to Accommodate the (x,y)DNA Bases

Alignment of the ceiling region of the finger domain (Val32, Ala42, Ala44, Asn47, Gly58, and Met76) that spans the dN:dNTP base pair is important for replication, and may be one factor that prevents base-pair mismatches [64,77,85]. For (x,y)DNA replication, the ceiling residues undergo some movement relative to the natural complexes (Figure S9). Specifically, for all (x,y)DNA bases, Met76 shifts away from the dN:dNTP pair by on average 0.7–0.9 Å (Table 6). Additionally, Val32 rotates towards dN:dNTP for dxG, dyG, dxT and dxC. Gly58 moves towards the dN:dNTP by ~1.0–1.3 Å for dyC, dxT and dxG, stabilized by a hydrogen bond between the template n+1 dC amine group and the Gly58 backbone carboxylate. For replication of the remaining modified bases, Gly58 moves away from the dN:dNTP pair by on average 0.8–1.2 Å.

### 2.8. Accommodation of the Modified Bases Disrupts the Enzyme Backbone, Particularly for the Finger and Little Finger Domains

To determine whether accommodation of d(x,y)N causes any significant changes in the overall Dpo4 ternary complex, beyond displacement of the template strand and the changes to the Dpo4:DNA hydrogen-bonding contacts previously discussed, the global movements of each domain were systematically analyzed. The RMSD per domain over each simulation (calculated relative to the corresponding representative structure) indicates that each domain exhibits nearly the same stability throughout the trajectory (Table S9). When the global structure of the d(x,y)N complexes are compared to the natural complexes, a repositioning of the protein backbone of the finger, thumb and little finger domains, as well as the tether region is observed and reemphasizes the structural conservation of the palm domain (Figure 7 and Figure S10). Furthermore, the smallest deviations from the natural systems occur for the dyT complex (Figure 7a), while the largest deviations occur for dxC (Figure 7b). These deviations are present throughout the simulation trajectories (Table S10), albeit slightly diminished when the complete modified simulation trajectories rather than single representative structures are compared to the corresponding natural representative structure. Despite changes in the enzyme–DNA hydrogen-bonding contacts in the finger region previously discussed, the observed repositioning does not significantly affect enzyme–DNA interactions in any other domain (Tables S3, S7, S8, and S11–S15).



**Figure 7.** Overlay of the C $\alpha$  protein backbone in representative MD structures of the Dpo4 ternary complex for the replication of: (a) dyT (blue); or (b) dxC (yellow), and dT (white). See Figure S10 for all systems.

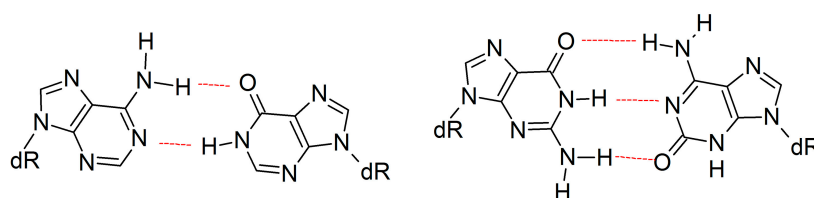
### 3. Discussion

Our molecular dynamics simulations provide important structural information for the experimentally observed replication of expanded (x) and widened (y) DNA nucleobases by the model bypass polymerase (Dpo4). Since previous literature proposed that Dpo4 replication of DNA containing base pairs with increased C1'–C1' distances may disrupt the alignment of the catalytic center [64], we initially focused on the structure of the active site relative to canonical DNA replication. The efficiency of the catalytic step is known to depend on discrete interactions between several Dpo4 residues and the dNTP backbone, as well as the coordination of two Mg<sup>2+</sup> ions [76]. Although the coordination of the Mg<sup>2+</sup> ions was unchanged relative to the natural complexes and the majority of enzyme–DNA hydrogen-bonding interactions are maintained, some significant differences in the interactions with the active site residues were observed. For example, the dNTP hydrogen-bonding interactions with Arg51, Thr45 and Tyr10 were disrupted for the replication of dxC and dyC, while Lys159 shifted in the ternary replication complexes for dyA, dxG, dyG, dxT and dyT. However, the changes observed for active site residues relative to natural DNA replication were not consistent for all modified bases. Another factor that may influence replication is the geometry of the newly forming base pair. Although the Watson–Crick hydrogen-bonding pattern was maintained for all modified bases, the geometry of the pair significantly fluctuates for dyA and dyC, which decreases the canonical interaction energy by up to 25 kcal·mol<sup>−1</sup>. Furthermore, while the overall base-step parameters are not considerably altered, there is a significant deviation in at least one base-pair parameter for each modified base (particularly buckle, propeller, opening or stretch), with the deviations typically being larger for the widened bases. Regardless of these differences, the reaction (distance and angle) parameters for dNTP insertion opposite expanded and widened bases are unchanged relative to natural DNA replication. Conservation of these reaction parameters correlates with the experimentally observed ability of Dpo4 to replicate (x,y)DNA [30,59]. Furthermore, some of the disruptions observed relative to the natural systems in our calculations may at least in part explain the up to 3,600 times less efficient replication of (x,y)DNA compared to the replication of natural DNA [30,59]. Additionally, there are notable correlations between trends in the experimentally observed reduction in replication rates and MD structural information for the ternary complex. Specifically, among the widened bases, dyT causes the least structural distortion to the insertion complex and correspondingly has the smallest reduction in the experimentally observed replication rate among the yDNA bases [59]. In contrast, the dxC insertion complex has the largest calculated distortion among the expanded bases, which correlates with the greatest reduction in replication rate among the xDNA bases [30]. Nevertheless, not all experimental observations for the replication of d(x,y)N can be correlated with the results obtained from MD simulations of the Dpo4 insertion complex. Indeed many factors govern the

experimentally observed decrease in rate, such as reaction barriers, pyrophosphate departure and/or pre/post-catalytic conformational changes, which cannot be accounted for in simulations of ternary replication complexes.

In order to maintain both the geometry of the active site and Watson–Crick hydrogen bonding, while accommodating a dN:dNTP pair with an increased C1'–C1' distance, Dpo4 allows the template backbone to shift away from the reaction center by up to 3 Å. This movement has direct impact on contacts between the template strand and Dpo4, including interactions with Lys78, Arg331, Ser34 and Ser40. Furthermore, interactions between the dN:dNTP pair and the ceiling region are affected by the accommodation of the (x,y)DNA base, particularly those involving Val32, Gly58 and Met76. Indeed, when representative structures of Dpo4 bound to DNA helices containing the modified bases are compared to those for the canonical counterparts (Figure 7 and Figure S10), the largest enzyme deviations are seen in the (finger and little finger) domains that surround the template strand, while there are relatively small deviations in the (palm) domain that contains the catalytic center. This reveals that it is the flexibility of the finger and little finger domains that allows Dpo4 to accommodate size-extended bases, without compromising the interactions required to stabilize and position the dNTP for catalytic nucleotide insertion. This flexibility may also explain the experimentally observed reduction in fidelity for Dpo4 replication of expanded bases, as well as the ability of Dpo4 to extend past (x,y)DNA, which was not possible for the more rigid Kf(*exo*–) polymerase [30,56,59]. Our conclusions correlate with previous computational studies of Dpo4 preinsertion, insertion and extension complexes for natural DNA replication, which conjecture that the finger and little finger domains are the most flexible regions and therefore may aid accommodation of modified DNA [81].

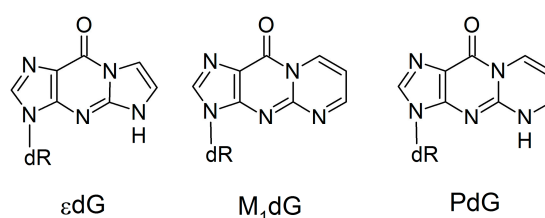
The present work can be used to gain insight into the replication of other modified bases that form complementary hydrogen-bonded base pairs with increased C1'–C1' distances. For example, *anti*-purine:*anti*-purine mismatches (e.g., adenine:hypoxanthine and guanine:isoguanine; Figure 8) are of great interest due to their relation to the origin of life and developing an understanding of why genetic information is stored in specific purine:pyrimidine base pairs [92–96]. Although *anti*-purine:*anti*-purine mismatches have been shown to form DNA helices with a stability that is similar to analogous purine:pyrimidine helices [92–94], the replication of such purine:purine pairs has not been studied to the best of our knowledge. Nevertheless, it has been speculated that the increased C1'–C1' distance will impede replication by misaligning the Dpo4 active center, particularly in the area surrounding the catalytic Mg<sup>2+</sup> ions [64]. However, this proposal directly contrasts our current observations for the replication of an expanded or widened DNA base. Therefore, we instead propose that movements in the finger and little finger domains of Dpo4 will accommodate DNA containing *anti*-purine:*anti*-purine mismatches, while maintaining the active center conformation and allowing for replication.



**Figure 8.** Examples of size-extended *anti*-purine:*anti*-purine mismatches.

Unlike *anti*-purine:*anti*-purine mismatches, experimental studies have investigated the Dpo4 replication of damaged DNA bases that would lead to pairs with increased C1'–C1' distances, such as 1,N<sup>2</sup>-ethenoguanine ( $\epsilon$ dG), malondialdehyde-guanine (M<sub>1</sub>dG) and 1,N<sup>2</sup>-propanoguanine (PdG; Figure 9) [97–99]. Similar to the (x,y)DNA bases, these damaged bases have expanded ring systems; however, the damaged bases lack the Watson–Crick hydrogen-bond face of dG. According to published crystal structures [97–99], these lesions do not significantly change the overall enzyme

structure. However, the incoming dNTP is not accommodated opposite the expanded bases, which instead result in 1-base deletion mutations. Since 1-base deletion mutations are formed because the dNTP pairs opposite the  $n+1$  base, the reaction parameters for modified base replication are misaligned and therefore there is a low replication efficiency (up to  $2 \times 10^4$  times slower than canonical DNA replication) [97–99]. This contrasts the observed lack of deletion mutations and successful replication of  $(x,y)$ DNA [30,59], which our work reveals maintains canonical Watson-Crick pairing within the Dpo4 active site through movement of the template DNA strand and surrounding enzyme domains. Comparison of these two general classes of modified DNA highlights the important role that hydrogen bonding plays in the replication of size-extended bases by Dpo4. Specifically, dN:dNTP hydrogen bonding can act as a driving force to move the template strand rather than allowing size-extended bases to disrupt the active site configuration. Interestingly, this conclusion contrasts previous assertions for the replication of modified bases that form pairs with the same shape (and therefore size) as canonical pairs, which do not have a strict requirement of dN:dNTP hydrogen bonding for nucleotide insertion [85,86,100,101].



**Figure 9.** Damaged DNA bases with extended ring systems that lack Watson–Crick hydrogen bonding.

#### 4. Materials and Methods

Starting structures for the MD simulations were obtained from a crystal structure of the Dpo4 ternary complex for the insertion of dATP opposite dT during extension past the benzo[a]pyrene adenine adduct (PDB ID: 1S0M, chains B, E and F) [102]. The catalytic divalent ions (Ca403 and Ca404) were modified to  $Mg^{2+}$  ions. Two excess divalent ions (Ca405 and Ca408) are crystallization artifacts and were removed. The template DNA strand sequence was modified to 5′-CNCCATCGCC (Figure S11), with dxN (dxA, dxC, dxG, or dxT) or dyN (dyA, dyC, dyG, or dyT; Figure 1) positioned at N. A complementary 8-mer primer DNA strand was paired opposite the 10-mer template strand and a complementary canonical dNTP nucleotide paired opposite dN. The inclusion of a single modified base in a canonical helix is consistent with previous experimental studies of the Dpo4 replication of  $(x,y)$ DNA [30,59]. To account for any differences in the polymerase active site upon purine or pyrimidine replication [77,78], two control simulations were performed by placing dG or dT at N paired opposite dCTP or dATP, respectively. The resulting 10 complexes were prepared for minimization using the tleap module of AmberTools 14 [103]. Specifically, hydrogen atoms were added to generate the natural protonation states of all DNA and protein residues, and the systems were neutralized with 5  $Na^+$  ions and solvated in an octahedral TIP3P water box such that the solute is at least 12.0 Å from the edge of the box. The resulting complexes each contain 341 amino acids, a dNTP, 18 nucleotides, 2  $Mg^{2+}$  ions, 5  $Na^+$  ions and ~15,300 water molecules.

The natural amino acids, nucleotides, and solvent were modeled using AMBER ff14SB parameters [104]. Ions were modeled using the 12-6-4 Lennard Jones-type non-bonded parameters of Li and Merz [105]. The dNTP, dxN and dyN parameters were assigned according to the GAFF [106] and AMBER ff14SB [104] force fields using the antechamber module of Amber 14 [103]. dNTP charges were taken from the literature [70,107,108]. Partial charges for dxN and dyN (Supplementary Tables S16–S23) were calculated using Gaussian 09 [109] (B3LYP/6-31+G(d)) and determined using RESP charge fitting with the PyRED (SEP-2015) server [109–113].

For all systems, the solvent was initially minimized and equilibrated using 1000 steps of steepest decent and 1000 steps of conjugate gradient minimization, followed by 10 ps of constant energy equilibration using a 0.001 ps time step. During solvent minimization and equilibration, a  $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  force constraint was placed on the protein, DNA and ions. Subsequently, the same steps were performed to minimize the solvent and  $\text{Na}^+$  ions by placing a  $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  force constraint on the protein, DNA and  $\text{Mg}^{2+}$  ions. Next, the Dpo4 complex was minimized using 5000 steps of steepest decent, followed by 5000 steps of conjugate gradient minimization, with a  $25 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  force constraint on the solvent and  $\text{Na}^+$  ions. The solvent and  $\text{Na}^+$  ions were then re-minimized using 100 steps of steepest decent, followed by 500 steps of conjugate gradient minimization, with a  $25 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  force constraint on the protein, DNA and  $\text{Mg}^{2+}$  ions. A non-bonded cutoff of 8.0  $\text{\AA}$  was implemented in all initial minimization and equilibration calculations.

Following initial minimization and equilibration calculations, the system was heated from 10.0 to 310.0 K over 10 ps using a 0.001 ps time step and the weak-coupling algorithm with a time constant of 2.0 ps, and a  $25 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  force constraint on the protein, DNA and  $\text{Mg}^{2+}$  ions. To ensure the systems reached the desired temperature of 310.0 K, the heating step was repeated, this time heating the system from 275.0 to 310.0 K. Next, the force constraints on the protein, DNA and  $\text{Mg}^{2+}$  ions were slowly lifted while maintaining the temperature at 310.0 K using Langevin dynamics (collision frequency of  $3.0 \text{ ps}^{-1}$ ). Specifically, the restraints were lifted over four steps, with the restraints set to 20, 5, 2.5 and 1.5  $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ , and run for 10, 30, 20, and 20 ps in each step, respectively. These calculations invoked a time step of 0.002 ps, the SHAKE algorithm and periodic boundary conditions, and maintained the pressure of the system at 1.0 bar via isotropic position scaling (pressure relation time of 1.0 ps). In the final equilibration, 2 ns of unrestrained simulation were performed with a 0.002 ps time step, SHAKE and periodic boundary conditions, the temperature maintained at 310.0 K through Langevin dynamics (collision frequency of  $3.0 \text{ ps}^{-1}$ ), the pressure maintained at 1.0 bar through isotropic position scaling (pressure relation time of 1.0 ps), and a non-bonded cutoff of 10.0  $\text{\AA}$  implemented.

Following minimization and equilibration, a 40 ns unrestrained production MD simulation with a 0.002 ps time step was performed on each of the 10 systems. During the production simulation, isotropic position scaling (pressure relation time of 2.0 ps) was used to maintain the pressure of the system at 1.0 bar, and the temperature of the system was held constant at 310.0 K through Langevin dynamics (collision frequency of  $3.0 \text{ ps}^{-1}$ ). SHAKE and periodic boundary conditions were implemented for production runs, with a non-bonded cutoff of 10.0  $\text{\AA}$ . Simulations were performed using the pmemd module of Amber 14 [103]. Each simulation remained stable, with a backbone RMSD of 1.1–1.6  $\text{\AA}$  over the 40 ns simulation (Table S24).

Base-step parameters were evaluated using Curves+ [114], while all other structural and energetic analyses was completed using AmberTools 14 [103]. For the energetic analysis of enzyme–DNA interactions, Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) pairwise energies were calculated for binding of the helix (including dNTP) to Dpo4. Furthermore, in line with previous reports [90], linear interaction energies were used to evaluate dN:dNTP hydrogen-bonding (electrostatic component) and stacking energies (van der Waals component) energies.

## 5. Conclusions

We have provided the first structural information for the replication of an expanded or widened DNA base by Dpo4. Our calculations reveal that an ( $x,y$ )DNA base paired opposite a canonical dNTP maintains Watson–Crick hydrogen bonding within the Dpo4 active site, albeit with some deviations to the base-pair and base-step parameters. The Dpo4 active site accommodates the increased  $\text{C1}'\text{--}\text{C1}'$  distance in these base pairs by allowing the template DNA strand to displace relative to the strand position during the replication of a natural base. As a result, the majority of hydrogen-bonding interactions between the enzyme and the incoming dNTP are maintained, while some deviations in

the hydrogen-bonding contacts with the template strand occur. This allows the reaction parameters to remain well aligned for dNTP insertion, which correlates with the experimentally observed successful replication of (x,y)DNA bases [30,59]. Since we did not observe consistent changes in the modified ternary complexes across the expanded and widened bases, further studies are required to fully explain the experimentally observed low Dpo4 replication efficiency for the (x,y)DNA bases [30,59].

In addition to providing insight into how expanded and widened DNA bases are accommodated in the Dpo4 active site, this work increases our fundamental understanding of Dpo4 function. First, although it was previously proposed that *anti*-purine:*anti*-purine mismatches with complementary hydrogen bonding and increased C1'–C1' distances would misalign the catalytic center and thus hinder replication, our work suggests that size-extended base pairs can be accommodated by moving the template strand and associated Dpo4 domains rather than disrupting the reaction site. Second, our work highlights that dN:dNTP hydrogen bonding plays a critical role in the bypass of size-extended bases that do not have the same shape as the canonical counterparts. Although these proposals must be validated by future experimental studies investigating the replication of a range of modified nucleobases that result in size-extended base pairs, our work contributes to the growing body of literature geared towards understanding the function of Dpo4 and other bypass polymerases.

**Supplementary Materials:** Supplementary materials can be accessed at: <http://www.mdpi.com/1420-3049/21/7/822/s1>.

**Acknowledgments:** This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, 249598-07); Canada Research Chairs Program (950-228175); and Canada Foundation for Innovation (22770). K.A.W also thanks NSERC (Vanier), Alberta Innovates–Technology Futures, and the University of Lethbridge for student scholarships. L.A. acknowledges Alberta Innovates–Technology Futures for an iCore postdoctoral fellowship. Computational resources from the New Upscale Cluster for Lethbridge to Enable Innovative Chemistry (NUCLEIC), and from Westgrid and Compute/Calcul Canada, are greatly appreciated.

**Author Contributions:** S.D.W., K.A.W. and L.A. conceived and designed the experiments, and wrote the paper. L.A. and K.A.W. performed and analyzed the experiments. All authors read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

Ala	alanine
Arg	arginine
Asn	asparagine
dA	deoxyadenosine monophosphate
dC	deoxycytidine monophosphate
DNA	deoxyribose nucleic acid
dNTP	deoxyribose nucleoside triphosphate
Dpo4	DNA polymerase IV
dG	deoxyguanosine monophosphate
dN	DNA nucleotide
dT	deoxythymidine monophosphate
dxA	expanded deoxyadenosine monophosphate
dxC	expanded deoxycytidine monophosphate
dxG	expanded deoxyguanosine monophosphate
dxT	expanded deoxythymidine monophosphate
dyA	widened deoxyadenosine monophosphate
dyC	widened deoxycytidine monophosphate
dyG	widened deoxyguanosine monophosphate
dyT	widened deoxythymidine monophosphate
εdG	1,N <sup>2</sup> -ethenoguanine
Glu	glutamine
Gly	glycine
Kf( <i>exo</i> -)	Klenow Fragment( <i>exo</i> -)
Lys	lysine



M <sub>1</sub> dG	malondialdehyde-guanine
MD	molecular dynamics
Met	methionine
Mg	magnesium
MM/GBSA	Molecular Mechanics / Generalized Born Surface Area
PdG	1,N <sup>2</sup> -propanoguanine
Phe	phenylalanine
RMSD	root-mean-square deviation
Ser	serine
Thr	threonine
Tyr	tyrosine
Val	valine
xDNA	expanded deoxyribose nucleic acid
yDNA	widened deoxyribose nucleic acid

## References

- Webb, S.J. Bioinspired organic chemistry. *Annu. Rep. Prog. Chem. Sect. B Org. Chem.* **2006**, *102*, 377–400. [[CrossRef](#)]
- Cobb, A.J.A. Recent highlights in modified oligonucleotide chemistry. *Org. Biomol. Chem.* **2007**, *5*, 3260–3275. [[CrossRef](#)] [[PubMed](#)]
- Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D.W. PNA: Synthetic polyamide nucleic acids with unusual binding properties. *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 2796–2823. [[CrossRef](#)]
- Kaur, H.; Babu, B.R.; Maiti, S. Perspectives on chemistry and therapeutic applications of locked nucleic acid (LNA). *Chem. Rev.* **2007**, *107*, 4672–4697. [[CrossRef](#)] [[PubMed](#)]
- Pinheiro, V.B.; Holliger, P. The XNA world: Progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245–252. [[CrossRef](#)] [[PubMed](#)]
- Moran, S.; Ren, R.X.F.; Rumney, S.; Kool, E.T. Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057. [[CrossRef](#)] [[PubMed](#)]
- Egli, M. The steric hypothesis for DNA replication and fluorine hydrogen bonding revisited in light of structural data. *Acc. Chem. Res.* **2012**, *45*, 1237–1246. [[CrossRef](#)] [[PubMed](#)]
- Malyshev, D.A.; Dhami, K.; Quach, H.T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F.E. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12005–12010. [[CrossRef](#)] [[PubMed](#)]
- LaPointe, S.M.; Wetmore, S.D. Characterization of the binding interactions between a unique class of modified purines and the natural nucleobases. *Chem. Phys. Lett.* **2005**, *408*, 322–328. [[CrossRef](#)]
- Polak, M.; Seley, K.L.; Plavec, J. Conformational properties of shape modified nucleosides—fleximers. *J. Am. Chem. Soc.* **2004**, *126*, 8159–8166. [[CrossRef](#)] [[PubMed](#)]
- Seley, K.L.; Salim, S.; Zhang, L.; O’Daniel, P.I. “Molecular chameleons” design and synthesis of a second series of flexible nucleosides. *J. Org. Chem.* **2005**, *70*, 1612–1619. [[CrossRef](#)] [[PubMed](#)]
- Hirao, I.; Kimoto, M.; Yamashige, R. Natural versus artificial creation of base pairs in DNA: Origin of nucleobases from the perspectives of unnatural base pair studies. *Acc. Chem. Res.* **2012**, *45*, 2055–2065. [[CrossRef](#)] [[PubMed](#)]
- Miguel, F.-C.; Vincent, M.; Bobby, G.S. Benzo-homologated nucleobases in a nanotube-electrode set-up for DNA sequencing. *Nanotechnology* **2007**, *18*, 424019.
- Sun, H.; Ren, J.; Qu, X. Carbon nanomaterials and DNA: From molecular recognition to applications. *Acc. Chem. Res.* **2016**, *49*, 461–470. [[CrossRef](#)] [[PubMed](#)]
- Deleavey, G.F.; Damha, M.J. Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.* **2012**, *19*, 937–954. [[CrossRef](#)] [[PubMed](#)]
- Sahu, N.K.; Shilakari, G.; Nayak, A.; Kohli, D.V. Antisense technology: A selective tool for gene expression regulation and gene targeting. *Curr. Pharm. Biotechnol.* **2007**, *8*, 291–304. [[CrossRef](#)] [[PubMed](#)]
- Linko, V.; Ora, A.; Kostiaainen, M.A. DNA nanostructures as smart drug-delivery vehicles and molecular devices. *Trends Biotechnol.* **2015**, *33*, 586–594. [[CrossRef](#)] [[PubMed](#)]

18. Hean, J.; Crowther, C.; Ely, A.; Ul Islam, R.; Barichiev, S.; Bloom, K.; Weinberg, M.S.; van Otterlo, W.A.; de Koning, C.B.; Salazar, F.; et al. Inhibition of hepatitis B virus replication in vivo using lipoplexes containing altritol-modified antiviral siRNAs. *Artif. DNA PNA XNA* **2010**, *1*, 17–26. [[CrossRef](#)] [[PubMed](#)]
19. Marimani, M.; Hean, J.; Bloom, K.; Ely, A.; Arbutnot, P. Recent advances in developing nucleic acid-based HBV therapy. *Future Microbiol.* **2013**, *8*, 1489–1504. [[CrossRef](#)] [[PubMed](#)]
20. Li, L.; Degardin, M.; Lavergne, T.; Malyshev, D.A.; Dhimi, K.; Ordoukhanian, P.; Romesberg, F.E. Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J. Am. Chem. Soc.* **2014**, *136*, 826–829. [[CrossRef](#)] [[PubMed](#)]
21. Teller, C.; Willner, I. Functional nucleic acid nanostructures and DNA machines. *Curr. Opin. Biotechnol.* **2010**, *21*, 376–391. [[CrossRef](#)] [[PubMed](#)]
22. Schmidt, M. Xenobiology: A new form of life as the ultimate biosafety tool. *Bioessays* **2010**, *32*, 322–331. [[CrossRef](#)] [[PubMed](#)]
23. Yang, Z.; Chen, F.; Alvarado, J.B.; Benner, S.A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **2011**, *133*, 15105–15112. [[CrossRef](#)] [[PubMed](#)]
24. Wilhelmsson, L.M. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.* **2010**, *43*, 159–183. [[CrossRef](#)] [[PubMed](#)]
25. Stevenson, T.M.; Leonard, N.J. Defined dimensional alterations in enzyme substrates. Lin-naphthoadenine and lin-naphthoadenosine. *J. Org. Chem.* **1984**, *49*, 2158–2164. [[CrossRef](#)]
26. Leonard, N.J. Dimensional probes of enzyme-cobinding sites. *Acc. Chem. Res.* **1982**, *15*, 128–135. [[CrossRef](#)]
27. Leonard, N.J.; Scopes, D.I.C.; Van der Lijn, P.; Barrio, J.R. Dimensional probes of the enzyme binding sites of adenine nucleotides. Biological effects of widening the adenine ring by 2.4 Å. *Biochemistry* **1978**, *17*, 3677–3685. [[CrossRef](#)] [[PubMed](#)]
28. Lu, H.; He, K.; Kool, E.T.  $\gamma$ DNA: A new geometry for size-expanded base pairs. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 5834–5836. [[CrossRef](#)] [[PubMed](#)]
29. Lee, A.H.; Kool, E.T. Novel benzopyrimidines as widened analogues of DNA bases. *J. Org. Chem.* **2005**, *70*, 132–140. [[CrossRef](#)] [[PubMed](#)]
30. Lu, H.; Lynch, S.R.; Lee, A.H.; Kool, E.T. Structure and replication of  $\gamma$ DNA: A novel genetic set widened by benzo-homologation. *ChemBioChem* **2009**, *10*, 2530–2538. [[CrossRef](#)] [[PubMed](#)]
31. Lee, A.H.F.; Kool, E.T. A new four-base genetic helix,  $\gamma$ DNA, composed of widened benzopyrimidine-purine pairs. *J. Am. Chem. Soc.* **2005**, *127*, 3332–3338. [[CrossRef](#)] [[PubMed](#)]
32. Liu, H.; Gao, J.; Kool, E.T. Size-expanded analogues of dG and dC: Synthesis and pairing properties in DNA. *J. Org. Chem.* **2005**, *70*, 639–647. [[CrossRef](#)] [[PubMed](#)]
33. Krueger, A.T.; Lu, H.; Lee, A.H.F.; Kool, E.T. Synthesis and properties of size-expanded DNAs: Toward designed, functional genetic systems. *Acc. Chem. Res.* **2007**, *40*, 141–150. [[CrossRef](#)] [[PubMed](#)]
34. Liu, H.; Lynch, S.R.; Kool, E.T. Solution structure of  $\chi$ DNA: A paired genetic helix with increased diameter. *J. Am. Chem. Soc.* **2004**, *126*, 6900–6905. [[CrossRef](#)] [[PubMed](#)]
35. Liu, H.; Gao, J.; Kool, E.T. Helix-forming properties of size-expanded DNA, an alternative four-base genetic form. *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402. [[CrossRef](#)] [[PubMed](#)]
36. Krueger, A.T.; Kool, E.T. Fluorescence of size-expanded DNA bases: Reporting on DNA sequence and structure with an unnatural genetic set. *J. Am. Chem. Soc.* **2008**, *130*, 3989–3899. [[CrossRef](#)] [[PubMed](#)]
37. Winnacker, M.; Kool, E.T. Artificial genetic sets composed of size-expanded base pairs. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 12498–12508. [[CrossRef](#)] [[PubMed](#)]
38. McConnell, T.L.; Wetmore, S.D. How do size-expanded DNA nucleobases enhance duplex stability? Computational analysis of the hydrogen-bonding and stacking ability of  $\chi$ DNA bases. *J. Phys. Chem. B* **2007**, *111*, 2999–3009. [[CrossRef](#)] [[PubMed](#)]
39. Fuentes-Cabrera, M.; Zhao, X.; Kent, P.R.C.; Sumpter, B.G. Electronic structure of  $\chi$ DNA. *J. Phys. Chem. B* **2007**, *111*, 9057–9061. [[CrossRef](#)] [[PubMed](#)]
40. Migliore, A.; Corni, S.; Varsano, D.; Klein, M.L.; Di, F.R. First principles effective electronic couplings for hole transfer in natural and size-expanded DNA. *J. Phys. Chem. B* **2009**, *113*, 9402–9415. [[CrossRef](#)] [[PubMed](#)]
41. Blas, J.R.; Huertas, O.; Tabares, C.; Sumpter, B.G.; Fuentes-Cabrera, M.; Orozco, M.; Ordejon, P.; Luque, F.J. Structural, dynamical, and electronic transport properties of modified DNA duplexes containing size-expanded nucleobases. *J. Phys. Chem. A* **2011**, *115*, 11344–11354. [[CrossRef](#)] [[PubMed](#)]

42. Jarchow-Choy, S.K.; Krueger, A.T.; Liu, H.; Gao, J.; Kool, E.T. Fluorescent xDNA nucleotides as efficient substrates for a template-independent polymerase. *Nucleic Acids Res.* **2011**, *39*, 1586–1594. [[CrossRef](#)] [[PubMed](#)]
43. Fuentes-Cabrera, M.; Sumpter, B.G.; Lipkowski, P.; Wells, J.C. Size-expanded yDNA bases: An ab initio study. *J. Phys. Chem. B* **2006**, *110*, 6379–6384. [[CrossRef](#)] [[PubMed](#)]
44. Fuentes-Cabrera, M.; Sumpter, B.G.; Wells, J.C. Size-expanded DNA bases: An ab initio study of their structural and electronic properties. *J. Phys. Chem. B* **2005**, *109*, 21135–21139. [[CrossRef](#)] [[PubMed](#)]
45. Fuentes-Cabrera, M.; Lipkowski, P.; Huertas, O.; Sumpter, B.G.; Orozco, M.; Luque, F.J.; Wells, J.C.; Leszczynski, J. Aromaticity-induced changes in the electronic properties of size-expanded DNA bases: Case of xC. *Int. J. Quantum Chem.* **2006**, *106*, 2339–2346. [[CrossRef](#)]
46. Sundaralingam, M.; Ponnuswamy, P.K. Stability of DNA duplexes with Watson-Crick base pairs: A predicted model. *Biochemistry* **2004**, *43*, 16467–16476. [[CrossRef](#)] [[PubMed](#)]
47. Zhang, L.; Li, H.; Chen, X.; Cukier, R.I.; Bu, Y. Absorption and fluorescence emission spectroscopic characters of size-expanded yDNA bases and effect of deoxyribose and base pairing. *J. Phys. Chem. B* **2009**, *113*, 1173–1181. [[CrossRef](#)] [[PubMed](#)]
48. Krueger, A.T. *Part I: Fluorescence of Size-Expanded DNA Bases: Reporting on DNA Sequence and Structure with an Unnatural Genetic Set. Part II: Toward Replication of xDNA, a Size-Expanded, Unnatural Genetic System*; Stanford University: Stanford, CA, USA, 2009.
49. Gao, J.; Liu, H.; Kool, E.T. Assembly of the complete eight-base artificial genetic helix, xDNA, and its interaction with the natural genetic system. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 3118–3122. [[CrossRef](#)] [[PubMed](#)]
50. Sale, J.E.; Lehmann, A.R.; Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 141–152. [[CrossRef](#)] [[PubMed](#)]
51. Voet, D.; Voet, J.G. *Biochemistry*, 4th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011.
52. Fijalkowska, I.J.; Schaaper, R.M.; Jonczyk, P. DNA replication fidelity in *Escherichia coli*: A multi-DNA polymerase affair. *FEMS Microbiol. Rev.* **2012**, *36*, 1105–1121. [[CrossRef](#)] [[PubMed](#)]
53. Yu, H.; Zhang, S.; Dunn, M.R.; Chaput, J.C. An efficient and faithful in vitro replication system for threose nucleic acid. *J. Am. Chem. Soc.* **2013**, *135*, 3583–3591. [[CrossRef](#)] [[PubMed](#)]
54. Loakes, D.; Holliger, P. Polymerase engineering: Towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* **2009**, 4619–4631. [[CrossRef](#)] [[PubMed](#)]
55. Chen, T.; Hongdilokkul, N.; Liu, Z.; Adhikary, R.; Tsuen, S.S.; Romesberg, F.E. Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat. Chem.* **2016**, *8*, 556–562. [[CrossRef](#)] [[PubMed](#)]
56. Chelliserrykattil, J.; Lu, H.; Lee, A.H.; Kool, E.T. Polymerase amplification, cloning, and gene expression of benzo-homologous “yDNA” base pairs. *Chembiochem.* **2008**, *9*, 2976–2980. [[CrossRef](#)] [[PubMed](#)]
57. Krueger, A.T.; Peterson, L.W.; Chelliserry, J.; Kleinbaum, D.J.; Kool, E.T. Encoding phenotype in bacteria with an alternative genetic set. *J. Am. Chem. Soc.* **2011**, *133*, 18447–18451. [[CrossRef](#)] [[PubMed](#)]
58. Delaney, J.C.; Gao, J.; Liu, H.; Shrivastav, N.; Essigmann, J.M.; Kool, E.T. Efficient replication bypass of size-expanded DNA base pairs in bacterial cells. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4524–4527. [[CrossRef](#)] [[PubMed](#)]
59. Lu, H.; Krueger, A.T.; Gao, J.; Liu, H.; Kool, E.T. Toward a designed genetic system with biochemical function: Polymerase synthesis of single and multiple size-expanded DNA base pairs. *Org. Biomol. Chem.* **2010**, *8*, 2704–2710. [[CrossRef](#)] [[PubMed](#)]
60. Fiala, K.A.; Suo, Z. Pre-steady-state kinetic studies of the fidelity of *Sulfolobus solfataricus* P2 DNA polymerase IV. *Biochemistry* **2004**, *43*, 2106–2115. [[CrossRef](#)] [[PubMed](#)]
61. Mizukami, S.; Kim, T.W.; Helquist, S.A.; Kool, E.T. Varying DNA base-pair size in subangstrom increments: Evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* **2006**, *45*, 2772–2778. [[CrossRef](#)] [[PubMed](#)]
62. Morales, J.C.; Kool, E.T. Varied molecular interactions at the active sites of several DNA polymerases: Nonpolar nucleoside isosteres as probes. *J. Am. Chem. Soc.* **2000**, *122*, 1001–1007. [[CrossRef](#)] [[PubMed](#)]
63. Xu, P.; Oum, L.; Lee, Y.C.; Geacintov, N.E.; Broyde, S. Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a Y-family DNA polymerase. *Biochemistry* **2009**, *48*, 4677–4690. [[CrossRef](#)] [[PubMed](#)]

64. Rechkoblit, O.; Malinina, L.; Cheng, Y.; Kuryavyi, V.; Broyde, S.; Geacintov, N.E.; Patel, D.J. Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxog lesion. *PLoS Biol.* **2006**, *4*, e11. [[CrossRef](#)] [[PubMed](#)]
65. Perlow-Poehnelt, R.A.; Likhterov, I.; Scicchitano, D.A.; Geacintov, N.E.; Broyde, S. The spacious active site of a Y-family DNA polymerase facilitates promiscuous nucleotide incorporation opposite a bulky carcinogen-DNA adduct: Elucidating the structure-function relationship through experimental and computational approaches. *J. Biol. Chem.* **2004**, *279*, 36951–336961. [[CrossRef](#)] [[PubMed](#)]
66. Xu, P.; Oum, L.; Geacintov, N.E.; Broyde, S. Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a Y-family DNA polymerase: A 5'-slippage mechanism. *Biochemistry* **2008**, *47*, 2701–2709. [[CrossRef](#)] [[PubMed](#)]
67. Brenlla, A.; Markiewicz, R.P.; Rueda, D.; Romano, L.J. Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment. *Nucleic Acids Res.* **2014**, *42*, 2555–2563. [[CrossRef](#)] [[PubMed](#)]
68. Patra, A.; Banerjee, S.; Johnson Salyard, T.L.; Malik, C.K.; Christov, P.P.; Rizzo, C.J.; Stone, M.P.; Egli, M. Structural basis for error-free bypass of the 5-N-methylformamidopyrimidine-dG lesion by human DNA polymerase  $\eta$  and *Sulfolobus solfataricus* P2 polymerase IV. *J. Am. Chem. Soc.* **2015**, *137*, 7011–7014. [[CrossRef](#)] [[PubMed](#)]
69. Perlow-Poehnelt, R.A.; Likhterov, I.; Wang, L.; Scicchitano, D.A.; Geacintov, N.E.; Broyde, S. Increased flexibility enhances misincorporation: Temperature effects on nucleotide incorporation opposite a bulky carcinogen-DNA adduct by a Y-family DNA polymerase. *J. Biol. Chem.* **2007**, *282*, 1397–1408. [[CrossRef](#)] [[PubMed](#)]
70. Zhang, L.; Rechkoblit, O.; Wang, L.; Patel, D.J.; Shapiro, R.; Broyde, S. Mutagenic nucleotide incorporation and hindered translocation by a food carcinogen C8-dG adduct in *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): Modeling and dynamics studies. *Nucleic Acids Res.* **2006**, *34*, 3326–3337. [[CrossRef](#)] [[PubMed](#)]
71. Wang, L.; Broyde, S. A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson-Crick pairing in the *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4). *Nucleic Acids Res.* **2006**, *34*, 785–795. [[CrossRef](#)] [[PubMed](#)]
72. Chandani, S.; Loechler, E.L. Molecular modeling benzo[a]pyrene N2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4. *J. Mol. Graph. Model.* **2007**, *25*, 658–670. [[CrossRef](#)] [[PubMed](#)]
73. Xu, C.; Maxwell, B.A.; Brown, J.A.; Zhang, L.; Suo, Z. Global conformational dynamics of a Y-family DNA polymerase during catalysis. *PLoS Biol.* **2009**, *7*, e1000225. [[CrossRef](#)] [[PubMed](#)]
74. Nair, D.T.; Kottur, J.; Sharma, R. A rescue act: Translesion DNA synthesis past N2-deoxyguanosine adducts. *IUBMB Life* **2015**, *67*, 564–574. [[CrossRef](#)] [[PubMed](#)]
75. Wang, L.; Wu, M.; Yan, S.F.; Patel, D.J.; Geacintov, N.E.; Broyde, S. Accommodation of a 1S-(–)-benzo[c]phenanthrenyl-N<sup>6</sup>-dA adduct in the Y-family Dpo4 DNA polymerase active site: Structural insights through molecular dynamics simulations. *Chem. Res. Toxicol.* **2005**, *18*, 441–456. [[CrossRef](#)] [[PubMed](#)]
76. Wang, L.; Yu, X.; Hu, P.; Broyde, S.; Zhang, Y. A water-mediated and substrate-assisted catalytic mechanism for *Sulfolobus solfataricus* DNA polymerase IV. *J. Am. Chem. Soc.* **2007**, *129*, 4731–4737. [[CrossRef](#)] [[PubMed](#)]
77. Vaisman, A.; Ling, H.; Woodgate, R.; Yang, W. Fidelity of Dpo4: Effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J.* **2005**, *24*, 2957–2967. [[CrossRef](#)] [[PubMed](#)]
78. Gaur, V.; Vyas, R.; Fowler, J.D.; Efthimiopoulos, G.; Feng, J.Y.; Suo, Z. Structural and kinetic insights into binding and incorporation of L-nucleotide analogs by a Y-family DNA polymerase. *Nucleic Acids Res.* **2014**, *42*, 9984–9995. [[CrossRef](#)] [[PubMed](#)]
79. Walsh, J.M.; Parasuram, R.; Rajput, P.R.; Rozners, E.; Ondrechen, M.J.; Beuning, P.J. Effects of non-catalytic, distal amino acid residues on activity of *E. Coli* DinB (DNA polymerase IV). *Environ. Mol. Mutagen.* **2012**, *53*, 766–776. [[CrossRef](#)] [[PubMed](#)]
80. Jarosz, D.F.; Godoy, V.G.; Delaney, J.C.; Essigmann, J.M.; Walker, G.C. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature* **2006**, *439*, 225–228. [[CrossRef](#)] [[PubMed](#)]
81. Wang, Y.; Arora, K.; Schlick, T. Subtle but variable conformational rearrangements in the replication cycle of *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) may accommodate lesion bypass. *Protein Sci.* **2006**, *15*, 135–151. [[CrossRef](#)] [[PubMed](#)]

82. Wang, Y.; Schlick, T. Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc.* **2008**, *130*, 13240–13250. [[CrossRef](#)] [[PubMed](#)]
83. Washington, M.T.; Helquist, S.A.; Kool, E.T.; Prakash, L.; Prakash, S. Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase  $\eta$ . *Mol. Cell. Biol.* **2003**, *23*, 5107–5112. [[CrossRef](#)] [[PubMed](#)]
84. Wolfle, W.T.; Washington, M.T.; Kool, E.T.; Spratt, T.E.; Helquist, S.A.; Prakash, L.; Prakash, S. Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase  $\kappa$ . *Mol. Cell. Biol.* **2005**, *25*, 7137–7143. [[CrossRef](#)] [[PubMed](#)]
85. Kool, E.T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **2002**, *71*, 191–219. [[CrossRef](#)] [[PubMed](#)]
86. Kool, E.T.; Morales, J.C.; Guckian, K.M. Mimicking the structure and function of DNA: Insights into DNA stability and replication. *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 990–1009. [[CrossRef](#)]
87. Liljas, A.; Liljas, L.; Piškur, J.; Lindblom, G.; Nissen, P.; Kjeldgaard, M. *Textbook of Structural Biology*; World Scientific: Singapore, Singapore, 2009.
88. Lu, X.-J.; Olson, W.K. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **2008**, *3*, 1213–1227. [[CrossRef](#)] [[PubMed](#)]
89. Arora, N.; Jayaram, B. Energetics of base pairs in B-DNA in solution: An appraisal of potential functions and dielectric treatments. *J. Phys. Chem. B* **1998**, *102*, 6139–6144. [[CrossRef](#)]
90. Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F.J.; Orozco, M.; Hobza, P. Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem. Eur. J.* **2006**, *12*, 2854–2865. [[CrossRef](#)] [[PubMed](#)]
91. Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell* **2001**, *107*, 91–102. [[CrossRef](#)]
92. Battersby, T.R.; Albalos, M.; Friesenhahn, M.J. An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids. *Chem. Biol.* **2007**, *14*, 525–531. [[CrossRef](#)] [[PubMed](#)]
93. Buckley, R.; Enekwa, C.D.; Williams, L.D.; Hud, N.V. Molecular recognition of watson-crick-like purine-purine base pairs. *ChemBioChem* **2011**, *12*, 2155–2158. [[CrossRef](#)] [[PubMed](#)]
94. Heuberger, B.D.; Switzer, C. An alternative nucleobase code: Characterization of purine-purine DNA double helices bearing guanine-isoguanine and diaminopurine 7-deaza-xanthine base pairs. *Chembiochem.* **2008**, *9*, 2779–2783. [[CrossRef](#)] [[PubMed](#)]
95. Cafferty, B.J.; Hud, N.V. Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA. *Isr. J. Chem.* **2015**, *55*, 891–905. [[CrossRef](#)]
96. Wächtershäuser, G. An all-purine precursor of nucleic acids. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1134–1135. [[CrossRef](#)] [[PubMed](#)]
97. Zang, H.; Goodenough, A.K.; Choi, J.-Y.; Irimia, A.; Loukachevitch, L.V.; Kozekov, I.D.; Angel, K.C.; Rizzo, C.J.; Egli, M.; Guengerich, F.P. DNA adduct bypass polymerization by *Sulfolobus solfataricus* DNA polymerase Dpo4: Analysis and crystal structures of multiple base pair substitutions and frameshift products with the adducts 1,*N*<sup>2</sup>-ethenoguanine. *J. Biol. Chem.* **2005**, *280*, 29750–29764. [[CrossRef](#)] [[PubMed](#)]
98. Wang, Y.; Musser, S.K.; Saleh, S.; Marnett, L.J.; Egli, M.; Stone, M.P. Insertion of dntps opposite the 1,*N*<sup>2</sup>-propanodeoxyguanosine adduct by *Sulfolobus solfataricus* P2 DNA polymerase IV. *Biochemistry* **2008**, *47*, 7322–7334. [[CrossRef](#)] [[PubMed](#)]
99. Eoff, R.L.; Stafford, J.B.; Szekely, J.; Rizzo, C.J.; Egli, M.; Guengerich, F.P.; Marnett, L.J. Structural and functional analysis of *Sulfolobus solfataricus* Y-family DNA polymerase Dpo4-catalyzed bypass of the malondialdehyde-deoxyguanosine adduct. *Biochemistry* **2009**, *48*, 7079–7088. [[CrossRef](#)] [[PubMed](#)]
100. Gahlon, H.L.; Schweizer, W.B.; Sturla, S.J. Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc.* **2013**, *135*, 6384–6387. [[CrossRef](#)] [[PubMed](#)]
101. Gahlon, H.L.; Boby, M.L.; Sturla, S.J. O<sup>6</sup>-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem. Biol.* **2014**, *9*, 2807–2814. [[CrossRef](#)] [[PubMed](#)]
102. Ling, H.; Sayer, J.M.; Plosky, B.S.; Yagi, H.; Boudsocq, F.; Woodgate, R.; Jerina, D.M.; Yang, W. Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2265–2269. [[CrossRef](#)] [[PubMed](#)]

103. Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; T.E. Cheatham, I.; Darden, T.A.; Duke, R.E.; Gohlke, H.; et al. *Amber 14*; University of California: San Francisco, CA, USA, 2014.
104. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725. [[CrossRef](#)] [[PubMed](#)]
105. Li, P.; Merz, K.M. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.* **2014**, *10*, 289–297. [[CrossRef](#)] [[PubMed](#)]
106. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [[CrossRef](#)] [[PubMed](#)]
107. Perlow, R.A.; Broyde, S. Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.* **2002**, *322*, 291–309. [[CrossRef](#)]
108. Zhang, L.; Shapiro, R.; Broyde, S. Molecular dynamics of a food carcinogen-DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol.* **2005**, *18*, 1347–1363. [[CrossRef](#)] [[PubMed](#)]
109. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; et al. *Gaussian 09*; Revision D.01; Gaussian, Inc.: Wallingford, CT, USA, 2013.
110. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821–7839. [[CrossRef](#)] [[PubMed](#)]
111. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J.C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517. [[CrossRef](#)] [[PubMed](#)]
112. Wang, F.; Becker, J.-P.; Cieplak, P.; Dupradeau, F.-Y. *Red Python: Object Oriented Programming for Amber Force Fields*; American Chemical Society: Washington, DC, USA, 2014.
113. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280. [[CrossRef](#)]
114. Lavery, R.; Moakher, M.; Maddocks, J.H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929. [[CrossRef](#)] [[PubMed](#)]

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).