Review Article

Architecture of Y-Family DNA Polymerases Relevant to Translesion DNA Synthesis as Revealed in Structural and Molecular Modeling Studies

Sushil Chandani,[1] Christopher Jacobs,[2] and Edward L. Loechler[1]

[1] *Biology Department, Boston University, Boston, MA 02215, USA*
[2] *Graduate Program in Bioinformatics, Boston University, Boston, MA 02215, USA*

Correspondence should be addressed to Edward L. Loechler, loechler@bu.edu

DNA adducts, which block replicative DNA polymerases (DNAPs), are often bypassed by lesion-bypass DNAPs, which are mostly in the Y-Family. Y-Family DNAPs can do non-mutagenic or mutagenic dNTP insertion, and understanding this difference is important, because mutations transform normal into tumorigenic cells. Y-Family DNAP architecture that dictates mechanism, as revealed in structural and modeling studies, is considered. Steps from adduct blockage of replicative DNAPs, to bypass by a lesion-bypass DNAP, to resumption of synthesis by a replicative DNAP are described. Catalytic steps and protein conformational changes are considered. One adduct is analyzed in greater detail: the major benzo[a]pyrene adduct (B[a]P-N^2-dG), which is bypassed non-mutagenically (dCTP insertion) by Y-family DNAPs in the IV/κ-class and mutagenically (dATP insertion) by V/η-class Y-Family DNAPs. Important architectural differences between IV/κ-class versus V/η-class DNAPs are discussed, including insights gained by analyzing ~400 sequences each for bacterial DNAPs IV and V, along with sequences from eukaryotic DNAPs kappa, eta and iota. The little finger domains of Y-Family DNAPs do not show sequence conservation; however, their structures are remarkably similar due to the presence of a core of hydrophobic amino acids, whose exact identity is less important than the hydrophobic amino acid spacing.

1. Introduction

DNA damaging agents (genotoxins) cause mutations that initiate tumor formation, which makes sense given that tumor cells have mutations in key growth control genes that lead to improperly regulated cell growth [1, 2]. The steps leading to mutagenesis vary depending on the genotoxin, but the paradigm in Figure 1 illustrates many of the typical steps using one particularly well-studied chemical carcinogen benzo[a]pyrene [3–5]. At the apex of this process are DNA adducts, which, if they are not removed by DNA repair, usually block replicative DNA polymerases (DNAPs). To overcome such potentially lethal blockage, cells have DNAPs that do translesion synthesis (TLS) past these DNA lesions/adducts [6–22].

Cells possess many DNAPs; for example, human cells, yeast (*S. cerevisiae*) and *E. coli* have at least fifteen, eight and five, respectively, [6–22]. Most TLS-DNAPs are in the Y-Family [6–22], where humans have three template-directed members (hDNAPs η, ι, and κ), yeast has one (scDNAP η), and *E. coli* has two (ecDNAPs IV and V). Y-Family DNAPs have a conserved ~350 aa core, which includes the polymerase active site (representative references [23–40]). As with all DNA polymerases, Y-Family members resemble a right-hand with thumb, palm, and fingers domains, although their "stubby" fingers and thumb result in more solvent accessible surface around the template/dNTP binding pocket [19], which is undoubtedly the case to accommodate the bulky and/or deforming DNA adducts/lesions that protrude into these open spaces during bypass. Y-Family DNAPs grip DNA with an additional domain, which is usually called the "little finger domain" or the "polymerase-associated domain" (PAD) [23–25].

Y-Family DNAPs are found in all three domains of life, bacteria, archaea, and eukaryotes, which undoubtedly reflects the fact that all cells face the same issues when confronting the need to replicate past DNA damage. The pattern of TLS is often strikingly similar in different cell types. For example, human DNAP κ was originally discovered because its sequence closely resembles *E. coli* DNAP IV [41–43], and dNTP insertion opposite a variety of adducts/lesions is remarkably similar for the DNAP IV/κ pair (Table 1), suggesting they are functional orthologs (discussed in [44]). *E. coli* DNAP V and human DNAP η are also functional orthologs, based on their similarity of dNTP insertion opposite a variety of adducts/lesions (Table 1, [44]). Cases have been made that the IV/κ-class is present in cells to bypass endogenously generated N^2-dG adducts, and the V/η-class is present to bypass UV-induced photoproducts, as discussed below.

B-Family DNAPs can also be involved in TLS, such as DNAP II in *E. coli* and REV3 (the polymerase subunit of DNAP ζ), which is present in most eukaryotes [6, 7, 13–15]. B-family TLS-DNAPs are involved in a DNA repair process involving some interstrand DNA cross-links [45–49] and in TLS of some adduct/lesions (see below).

Herein, we reflect principally on how structural architecture of Y-family DNAPs might affect their mechanism as it relates to cellular function, in particular why lesion-bypass is sometimes nonmutagenic and other times it is mutagenic. Extensive reviews that focus more on the cell biology, regulation and phenomenology have appeared recently for Y-Family DNAPs from bacteria [21, 22] and eukaryotes [7–12].

2. Translesion Synthesis DNA Polymerases in *E. coli*

E. coli has proven to be an excellent model system to study many aspects of the bypass of DNA adducts/lesions by TLS-DNAPs. *E. coli* has two Y-Family DNA polymerases: DNAP IV (*dinB* gene, 351 aa, 39.5 kDa) and DNAP V, which consists of one subunit of UmuC (*umuC* gene, 422 aa, 47.7 kDa) and two subunits of UmuD' (see below). UmuD' is derived from UmuD (*umuD* gene, 139 aa, 15 kDa) following autodigestive removal of its 24 N-terminal aa, when stimulated by RecA* [13, 14, 21, 22, 50]. DNAP II (*polB* gene, 783 aa, 90 kDa) is a B-Family lesion bypass DNAP. DNAPs II, IV and V are each induced as part of the SOS response, which is triggered by DNA damage and leads to the induction of ~40 proteins that help *E. coli* cope with the damage [50]. The basal and SOS-induced levels are different for each polymerase, where the [uninduced/induced] levels are [~40/~280] for DNAP II, [~250/~2500] for DNAP IV and [~15/~200] for DNAP V [51–53]. It seems likely that each of these TLS-DNAP is present in *E. coli* principally to overcome the cellular problems presented by a lesion commonly encountered in cells as discussed next.

Although DNAP V replicates undamaged templates with relatively low fidelity (10^{-3} to 10^{-4}) [54], one striking quality is its ability to accurately bypass UV photoproducts; for example, it inserts dATP opposite TT-CPDs [54]. Analysis of insertion tendencies opposite a variety of adducts/lesions led to the observation that DNAP V may have two insertion modes: (i) correct dNTP insertion, and (ii) default dATP insertion [44]. UV light is a frequently encountered form of DNA damage for which a TLS-DNAP might be important, and since TT-CPDs are the major UV lesion [55], a default dATP insertion mode might help minimize UV mutagenesis. However, the utilization of this second mode in other circumstances may have drawbacks. For example, UV mutagenesis also depends on the *umuD/C* genes, implying that DNAP V is required for UV mutagenesis, where C → T mutations in 5'-Py$\underline{C}$ sequences predominate, which also implies dATP insertion (discussed in reference [56]). DNAP V is involved in other mutagenesis pathways; for example, it inserts dATP opposite +BP in the G → T mutational pathway [57], as discussed below. In fact, the preferential mutagenic insertion of dATP opposite a variety of DNA lesions in *E. coli* has been called the "A-rule" (see [58, 59] and references therein), and it seems likely that this is attributable to DNAP V's tendency to insert dATP [44]. Based on lesion-bypass specificity (Table 1), *E. coli* DNAP V appears to be the functional ortholog of human DNAP η [36], which is almost certainly responsible for correct bypass of UV-lesions in human cells and minimizing UV-light mutagenesis that leads to skin cancer [60–65].

On its own, UmuC, which is the polymerase subunit of DNAP V, either misfolds or aggregates and is found in inclusion bodies [22, 50, 66]. UmuC copurifies with UmuD', though the yield is invariably low [22, 50, 66]. RecA is also required for efficient DNAP V activity, and recently, the "DNAP V mutasome" was shown to be a UmuC/UmuD'$_2$/RecA heterotetramer [67, 68]. The RecA monomer is added from the 3'-end of a RecA filament either in *cis* or in *trans*, where the former seems intuitively more likely, since UmuC/UmuD'$_2$ would encounter a 3'-*cis*-RecA at a lesion site, given that RecA filaments coat ss-DNA on the downstream side of a lesion-blocked replication fork. To form RecA filaments on ss-DNA, SSB must first be removed, which is accomplished by RecFOR [69]. Interestingly, some evidence suggests that the RecA eukaryotic homolog Rad51 is able to stimulate DNAP η, which is the DNAP V ortholog [70]. β-clamp also plays a significant role with DNAP V as discussed below.

DNAP IV replicates undamaged DNA only ~5-fold less accurately than the catalytic α-subunit of DNAP III [54]. It is prone to making −1 frameshift mutations in homopolymeric runs of six or more G : C base pairs, and base substitutions also result [22, 71]. DNAP IV's most striking quality is its ability to accurately bypass a variety of N^2-dG adducts [72–77]. Methylglyoxal is produced nonenzymatically from various cellular trioses and forms N^2-(1-carboxyethyl)-2'-dG as its major stable adduct, which is bypassed accurately by DNAP IV [76]. Oxidative metabolism forms reactive oxygen species that generate lipid peroxidation products that give exocyclic adducts, some of which can ring-open to N^2-dG adducts in ds-DNA [78] and might be bypassed by DNAP IV, though this has not been investigated experimentally. These observations have led several groups to speculate that the cellular rationale for the genesis of the IV/κ-class of Y-Family

TABLE 1: Dominant dNTP insertions opposite various DNA adducts/lesions by *E. coli* DNAPs IV and V, and human DNAPs κ and η.[†]

Lesion	DNAP V	DNAP η	DNAP IV	DNAP κ
[+ta]-BP-N^2-dG	A/C	A $\geq$ G	C	C
AAF-C8-dG	C	C	C/T	C/T
AF-C8-dG	—	—	C	C
TT-CPD	AA	AA	n	n
T(6-4)T	AG	nG	n	n
AP site	A	A	n	A*

[†]Dominant dNTP insertion using purified DNAPs, where "n" indicates "no" or low activity, "A*" indicates bypass by an unusual mechanism, and "—" indicates data unavailable. Data, as reviewed in [44].

DNAPs is the accurate bypass of N^2-dG adducts derived from various endogenous mechanisms [72, 76].

No analogous story vis-a-vis adducts/lesions has yet emerged to provide a rationale for the presence of B-family DNAP II in cells, though one possibility is its involvement in an accurate DNA repair pathway for interstrand cross-links [45]. An analogous pathway involving B-family DNAP ξ exists in eukaryotic cells, and a pathway has been proposed [46–49]. As discussed below, DNAP II functions in other TLS pathways.

UmuD$_2$C (not UmuD'$_2$C) is thought to slow down normal DNA replication in response to DNA damage, thus allowing additional time for lesion removal, which is considered a DNA damage checkpoint analogous to what happens in eukaryotic cells [79]. Another mechanism to accomplish this was recently described: DNAP II or IV can associate with the DnaB helicase and slow down the replication fork [80].

The TLS-DNAPs also confer selective advantage on *E. coli* during long periods in stationary phase, the so-called "growth advantage in stationary phase" (GASP) phenotype [81]. Finally, DNAP IV is particularly elevated in stationary phase ($\sim$7500/cell) and is implicated in adaptive mutagenesis [82].

3. Eukaryotic Y-Family DNAPs

Extensive reviews of the cell biology, regulation, and phenomenology of eukaryotic Y-Family DNAPs have appeared recently [7–12]. Herein, we focus on structural considerations that relate to lesion bypass, though we briefly describe each of the four subclasses of eukaryotic Y-Family DNAPs: REV1, DNAP κ, DNAP η, and DNAP ι.

REV1 is not a traditional template-directed polymerase and does not use base-base hydrogen bonding. Rather REV1 is a dCTP insertase that flips template dGs out of the helix, after which dCTP insertion is directed by hydrogen bonding to a REV1 arginine residue [83]. REV1 seems to play a central role in many lesion bypass events as a structural component, and DNAPs κ, η, and ι each have REV1 binding domains [9].

DNAP κ is the eukaryotic ortholog of DNAP IV (Table 1), and they seem to be present in cells to accurately insert opposite N^2-dG adducts [84], for example, DNAP κ deficient cells are sensitized to killing by benzo[a]pyrene, which

predominantly forms an N^2-dG adducts [85]. DNAP κ uniquely has an N-terminal extension of 100 aa called the "N-clasp" [35]. The N-clasp has three α-helices in a U-shape, one of which ($\sim$aa30–50) binds on the surface of the fingers domain, the second of which ($\sim$aa50–75) links the fingers and thumb domains and lies diagonally across the duplex region of DNA, and the third traverses the thumb domain to the usual site of the N-terminus in Y-Family DNAPs. Removal of the N-clasp significantly decreases DNAP κ polymerase activity [35]. The presence of the N-clasp has implications for lesion bypass; for example, DNAP κ does not bypass the N^6-dA adduct of benzo[a]pyrene [86], which has been attributed to a steric clash between the N-clasp and the pyrene moiety, as revealed in a molecular modeling study [87]. DNAP κ structure is considered below.

A major role of DNAP η is nonmutagenic bypass of UV lesions, such as TT-CPDs, and humans deficient in DNAP η have the cancer-prone syndrome Xeroderma pigmentosum variant (XPV), which leads to a high incidence of UV-induced skin cancer [60–65]. Both human and yeast DNAP η preferentially insert dATP opposite the 5$'$-T and 3$'$-T of a TT-CPD, with misinsertion being higher at the 3$'$-T, where dGTP is incorporated $\sim$3% of the time [88, 89]. Recently, X-ray structures of a TT-CPD in the active site of yeast DNAP η [39] and human DNAP η [40] have emerged. These findings are presented in a separate section(Section 8), after certain principles about Y-Family DNAPs structure have been discussed. DNAP η also plays a role in accurate bypass of the oxidative lesion 8oxoG [90] and adducts formed by the anticancer drug *cis*-platinum [91].

The role of DNAP ι in cells is more enigmatic, though the fact that deficient cells show enhanced sensitivity to oxidative damage may be revealing [92]. One interesting feature of DNAP ι is its propensity to use *syn*-purines in the template to form Hoogsteen base pairs *syn*-A : T and *syn*-G : C[36, 37]. A cellular rationale for this is the following. Oxidative damage leads to lipid peroxidation products, which form exocyclic adducts that block the Watson-Crick *anti*face of the DNA bases. Exocyclic purine adducts in the *syn*-configuration can still base pair through their Hoogsteen face. The *syn*-A : T base pair has two hydrogen bonds, just like *anti*-A : T. However, both *syn*-G : C and *syn*-G : T base pairs have only one hydrogen bond, unless a proton is trapped. In a *syn*-GH$^+$: C base pair the trapped proton is between N7G and N3C, whose pKa values are relatively
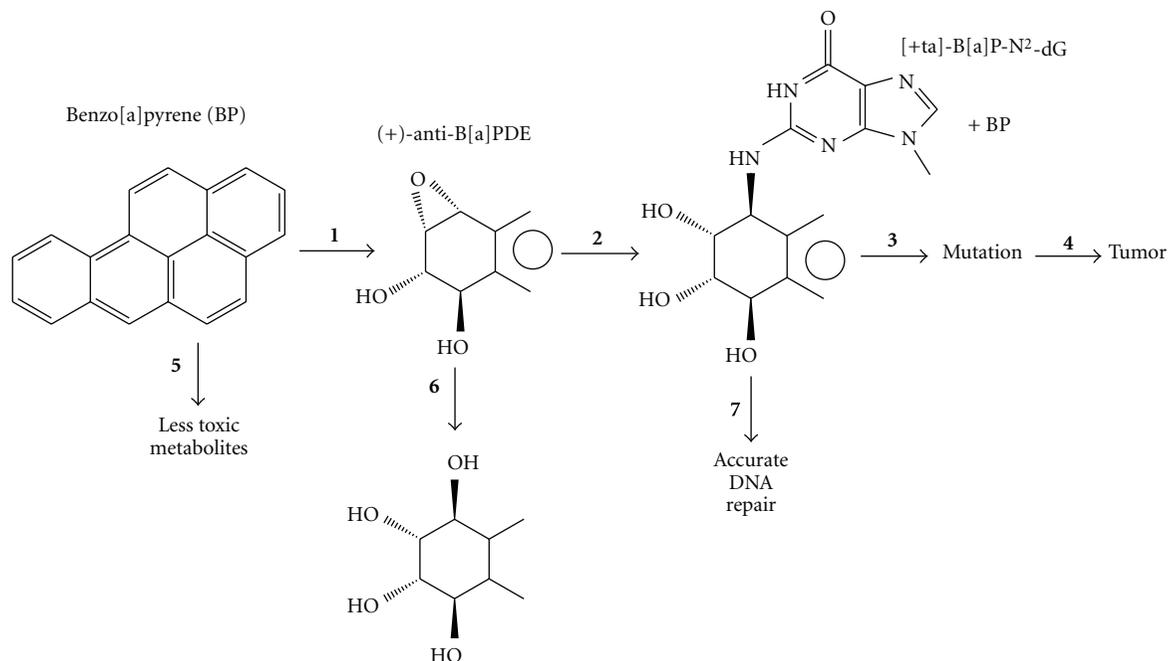
FIGURE 1: Mutagenesis/carcinogenesis paradigm with benzo[a]pyrene. Steps in the horizontal direction lead toward carcinogenicity, and include metabolic activation (step 1), reaction with DNA (step 2), adduct mutagenesis (step 3), and tumorigenesis (step 4). Steps in the vertical direction lead to diminished carcinogenicity, and include metabolic detoxification (step 5), carcinogen deactivation (step 6), and DNA repair (step 7). Diminished carcinogenicity is also associated with other cellular processes, such as delaying the cell cycle and apoptosis [1].

high ($\sim$3), while a trapped proton in a *syn*-GH$^+$: T base pair would be between O^{6}G and O^{4}T, whose pKa values are much lower. Thus, a *syn*-GH$^+$: T base pair is expected to be less stable than a *syn*-GH$^+$: C base. Evidence for DNAP ι using *syn*-GH$^+$: C base pairing exists [37]. 1,N^6-etheno-A directs both dCTP and dTTP incorporation, and in this case both require a trapped proton [38], where the pKa values of the relevant atoms trapping the proton are more equal. While such thinking is considered satisfying [12], one finding suggests that the situation can be more complex. DNAP ι preferentially incorporates dCTP opposite the major adduct of 2-acetylaminofluorene (AAF-C8-dG) [93], *Syn*-AAF-C8-dG places the bulky AAF-moiety in the minor groove, where molecular modeling showed that it does not fit, while *anti*-AAF-C8-dG : C pairing, which places the AAF-moiety in the spacious major groove side of DNAP ι, is possible [94]. The authors propose that purine adducts with bulk on the minor groove side probably use *syn*-purine pairing, but that purine adducts with bulk on the major groove side probably use *anti*purine pairing. *Anti*-AAF-C8-dG : C pairing requires a modest change in sugar pucker (from C3$'$-endo to C1$'$-exo), as noted in modeling studies with both DNAP ι [94] and Dpo4 [95].

Base substitution rates on undamaged templates are relatively high with all of these polymerases: yDNAP η ($\sim$ 10^{-2}), hDNAP η ($\sim$3.5 $\times$ 10^{-3}), h DNAP κ ($\sim$6 $\times$ 10^{-4}), and hDNAP ι actually prefers to form template-dT : dGTP; indel mutation rates are all in the same range ($\sim$1–2.4 $\times$ 10^{-3}) (reviewed in [7]).

4. Y-Family DNAP Mechanistic Steps

A number of comprehensive reviews have appeared that analyze the structures of Y-Family DNAPs [10–12, 32, 33]. In this section, we focus on what is known about protein structural changes that occur during DNA synthesis as probed via X-ray structural analysis and other techniques, principally with Dpo4. The chemistry of catalysis is also considered.

Upon DNA binding to Apo-Dpo4, the thumb/palm/ fingers domains do not change their structure dramatically. However, the little finger domain acts like a door, which is open in Apo-Dpo4, and then rotates $\sim$130° to close around DNA; in particular, it binds in the major groove in the duplex region from about L + 3 to L + 8 [33]. This motion is facilitated by the fact that the little finger is connected to the rest of the protein by a simple ten amino acid tether. Once binary-Dpo4 is formed, the palm, fingers and little finger translate $\sim$3.3 Å along the helix as the next template base slides into the active site, which opens the space into which the complementary dNTP binds to give ternary-Dpo4 [32]. The thumb domain, however, does not move in this step, but, rather, moves either before, during, or after the subsequent covalent reaction step. A variety of subtler changes in Dpo4 structure are also reported to accompany these steps [32, 33]. Kinetic studies reveal that Y-Family DNAPs have a rate-determining conformational change before dNTP incorporation [10, 96], and three conformational states E, E$'$, and E$''$ have been reported, where the E$'$ → E$''$ conformational transition is rate

determining, though the nature of these states have not been identified. Recently, hydrogen-deuterium exchange in tandem with mass spectrometry has been used to study conformational changes in Dpo4 brought about by dNTP binding [30]. Correct dNTP binding affects the structure of a loop between the B-helix and the C-helix above the Dpo4 active site. (The positioning of these features can be inferred from the UmuC(V) sequence in Figure 2.) Another conformational change was also detected in the H-helix, which contacts the primer strand and was proposed to move away from the active site in conjunction with ds-DNA movement to permit room for correct dNTP binding. The F-helix also moves, but this motion is not specific for the correct dNTP. In terms of lesion bypass, Dpo4 showed decreased catalytic efficiency with increasing bulk of N^2-dG adducts, which was attributed more to the effects of the bulky lesion on the rate of the catalytic step than on the rate of the conformational steps [31].

Several studies have shown that dNTP incorporation is more dependent on base:base hydrogen bonding for Y-Family DNAPs than for DNAPs in other families during the replication of both undamaged and damaged DNA [97–99].

The steps in covalent catalysis by Dpo4 have been explored using a combination of molecular modeling/dynamics and ab initio QM/MM minimizations; a novel water-mediated and substrate-assisted mechanism was proposed [100]. In the first step, a water molecule in the active site serves as a conduit to deprotonate the primer $3'$-OH and protonate an oxygen on the α-phosphate of the dNTP. In the second step, a second water molecule in the active site serves as a conduit to deprotonate the oxygen on the α-phosphate of the dNTP and to protonate an oxygen on the γ-phosphate. Following these two steps the deprotonated $3'$-O$^-$ of the primer is a stronger nucleophile and attacks the α-phosphate, while the second water molecule serves as a conduit again—this time to deprotonate the γ-phosphate of the dNTP and to protonate the β-phosphate, which is on the pyrophosphate leaving group, thus facilitating its removal.

5. The Steps Leading to Translesion Synthesis in *E. coli*

A well-developed model for the steps in translesion synthesis has emerged for *E. coli* [107]. Replicative DNAP III stalls at many adducts. For example, in the case of AAF-C8-dG, the $3' \rightarrow 5'$ exonuclease activity of DNAP III competes with its polymerase activity, such that [L-1]:[L0] ratio is ~10:1 ratio of primers, as determined in vitro [66]. A TLS-DNAP probably helps dissociate a stalled DNAP III from the lesion site (see below). DNAP III reinitiates replication hundreds to thousands of base pairs downstream of the adduct/lesion at the next primosome assembly site in a process called "replication restart," either on the lagging strand using the normal lagging strand machinery (i.e., PriA/B/C, DnaB/C/T, and primase), or on the leading strand, whose details are being worked out [108, 109]. This leaves an ss-gap between the lesion site and the site where DNAP III did replication restart. This gap is either filled via recombination or via DNA

replication, which begins with the action of TLS-DNAPs [15, 108, 109].

DNAP IV binds β-clamp to help release a stalled DNAP III from the same β-clamp, leaving DNAP IV/β-clamp at the site of the lesion [110]. This process is rapid ($t < 15$ s). Presumably, a similar mechanism operates for each TLS-DNAP (II, IV and V), which all have β-clamp binding sites (consensus: QLxLF) that are required for them to be active in *E. coli* [111]. An X-ray structure shows that the underlined amino acids Q̲L̲VL̲G̲L̲ at the C-terminus of DNAP IV form the main interactions with a "cleft" in the β-clamp [112]. The α-subunit of DNAP III and the δ-subunit of the γ-complex also bind to the cleft in the β-clamp. DNAP IV and V can also bind to a site in the "rim" of the β-clamp, but this seems unimportant for TLS [113–115]. In vitro studies show that β-clamp stimulates both polymerase activity and processivity of TLS-DNAPs: the addition of β-clamp in vitro increases DNAP IV activity ~2000-fold and processivity from 1 nucleotide to ~400 nucleotides, and also increases DNAP V activity ~100-fold and processivity from 1-2 nucleotides to ~18 nucleotides [22].

What factors affect the choice about which TLS-DNAP will insert opposite a particular lesion? Several lines of evidence suggest that *E. coli* has a hierarchy for the replication of normal, unadducted DNA when DNAP III is inactivated: DNAP II > IV > V [116]. (The assays did not permit an assessment of DNAP I.) Since this order (III > II > IV > V) does not reflect the relative concentration of these DNAPs in cells (see above), another mechanism for decision making was suggested, such as relative DNAP affinity for the β-clamp. This order does reflect relative fidelity of these DNAPs and would be a sensible order for *E. coli* to allow TLS-DNAPs to initially sample adducts/lesions prior to a decision about which will do TLS. But the ultimate decision is probably predominantly controlled by which TLS-DNAP is most efficient at bypassing a particular adduct/lesion biochemically.

After insertion opposite the lesion, additional extension synthesis by a TLS-DNAP is required, or else DNAP III's proof-reading $3' \rightarrow 5'$ exonuclease activity will remove the inserted nucleotides back to the site of the lesion [66, 117]. The amount of extension required before DNAP III can resume normal synthesis appears to be pathway dependent, where it is [L + 4] for the AAF-C8-dG nonmutagenic pathway with DNAP V, and [L + 3] for the AAF-C8-dG −2 frameshift pathway with DNAP V [66, 83, 117].

6. Two Case Studies Showing the Interplay of Y-Family DNAPs in Translesion Synthesis

More is known about the details of TLS for the major adduct of *N*-2-acetylaminofluorene (AAFC8-dG), and N^2-dG adducts in *E. coli* than for any other adducts/lesions in any other model system. In these cases, multiple translesion DNAPs are involved in both the nonmutagenic and mutagenic pathways, as outlined in this section.

AAF was originally developed as a potential pesticide, but it was abandoned when it was found to be a potent rat carcinogen [118]. Following activation, AAF principally binds at

Block 1 (residues 1–45)

```
Secondary   1        1        10        A        20        2    30        3  40        B 45
Consensus   M F A L C D V N N F Y A S C E R V F R P D L R G K P V V L S N N D G C V I A R S A E A K
Percent     55 56 84 87 42 100 31 87 43 88 97 57 95 81 97 46 57 86 54 97 46 85 29 54 40 89 67 70 92 91 93 94 94 94 95 84 54 55 84 74 72 28 74 91 89
AA type     h h u l l D h h s s h h S S s E p h h F P P D L h . h s h V L S N N D G s h l u t s . t A +
E. coli     M F A L C D V N A F Y A S C E T V F R P D L W G K P V V L S N N D G s h I A R N A E A K

A   34 0 346 5 40 0 58 8 35 6 0 236 13 47 0 9 41 1 5 6 41 0 47 3 22 7 23 4 8 1 1 2 1 7 5 53 2 346 45 0 116 10 372 1
C   1 0 4 5 55 0 129 0 9 20 0 66 0 334 1 14 1 0 1 0 0 0 15 0 1 0 4 2 1 0 0 1 1 1 386 0 0 344 21 5 1 19 7 0
D   3 0 0 0 0 411 0 19 0 0 0 0 0 0 0 0 0 2 76 0 190 0 1 42 0 0 0 0 1 0 1 1 4 386 0 0 0 1 0 1 0 24 13 0
E   1 0 0 0 0 0 0 0 0 0 0 0 0 0 397 9 0 0 3 0 33 0 38 4 13 0 0 0 0 0 0 0 0 0 0 2 1 0 0 0 22 305 0 0
F   3 229 0 1 0 0 0 0 0 363 13 0 0 0 0 0 2 353 0 0 1 2 1 3 1 0 0 0 0 0 6 7 4 4 8 392 0 0 6 0 0 1 1 0
G   0 0 28 0 0 0 43 0 14 0 0 1 0 0 0 0 1 0 0 0 0 0 222 0 2 1 4 0 6 7 0 2 1 0 0 0 4 0 0 22 0 1 1 0
H   0 1 0 21 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 9 0 3 9 6 0 0 0 1 1 0 1 0 1 0 0 1 0 0 0 0 2 0 0
I   23 72 3 14 113 0 31 0 0 0 0 1 0 0 2 0 1 21 10 0 0 16 16 0 5 0 78 87 23 0 0 0 0 0 0 0 11 103 227 0 2 0 5 1 0
K   7 0 0 0 0 0 0 15 2 0 0 0 0 0 0 0 44 0 2 5 0 35 0 61 19 166 0 0 1 3 0 1 0 1 2 1 0 0 2 0 22 4 0 366
L   11 11 11 358 10 0 11 0 0 0 1 0 0 0 0 2 78 29 0 1 0 351 8 0 6 10 36 6 373 0 0 0 5 5 15 0 7 0 0 26 11 0
M   224 0 8 2 7 0 4 0 22 0 0 1 0 1 2 8 6 0 0 1 3 0 0 3 0 0 0 0 6 0 10 0 0 0 5 0 0 2 1 0
N   2 0 0 0 0 0 2 358 175 0 0 0 1 0 0 1 0 0 74 0 5 0 34 82 2 1 0 0 5 385 385 2 0 16 0 0 88 62 0 0 0
P   9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 397 0 1 0 0 1 365 0 0 1 1 0 2 1 0 1 3 0 51 2 0 0
Q   1 0 0 3 1 0 1 2 2 0 0 0 0 0 0 6 79 0 0 18 0 9 0 15 3 12 0 0 1 1 0 2 1 0 1 0 0 3 0 11 37 0 1
R   1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 190 3 3 220 1 77 0 121 13 135 0 0 0 2 5 6 4 0 0 0 0 304 0 52 1 0 39
S   12 0 0 9 0 0 11 1 169 0 0 1 391 1 0 6 3 17 1 1 2 41 0 3 2 7 0 0 1 382 11 4 1 7 6 1 5 56 2 296 8 3 4 0
T   12 13 0 0 0 0 1 0 3 0 0 8 4 0 0 58 3 0 0 1 9 0 6 10 15 0 2 2 2 4 2 4 2 1 5 3 19 4 1 3 2
V   62 5 2 174 0 0 121 4 0 0 0 97 0 22 0 0 233 0 0 1 0 13 1 41 2 276 289 377 9 0 0 1 0 10 220 156 2 14 0 2 0 15 2
W   2 8 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 30 0 25 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Y   1 69 0 0 0 0 0 0 0 0 397 0 0 0 0 0 0 2 0 0 0 0 6 4 1 0 0 0 0 1 1 0 0 1 0 0 0 0 0 32 0 0 0
X   0 3 3 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 12 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 1 1 0

HYDRO   370 394 380 385 399 0 354 12 44 411 411 401 13 410 1 28 384 403 6 406 45 409 129 5 59 400 407 407 407 397 2 1 3 1 8 385 408 405 350 95 7 202 45 403 3
LONG    90 96 92 94 97 0 86 3 11 100 100 98 3 100 0 2 93 98 1 100 31 1 14 97 99 99 99 97 0 2 94 99 99 85 23 2 49 81 98 1
SHORT   3 0 1 0 0 1 92 44 0 0 0 0 0 0 0 75 21 36 10 0 99
SHORT   37 13 0 3 10 0 14 379 353 0 0 9 397 1 12 374 27 6 318 1 113 0 240 129 332 8 2 0 4 5 398 403 400 14 10 26 3 6 61 314 403 159 46 7 408
(lower)  9 3 0 1 0 2 3 92 86 0 0 2 97 0 2 91 7 1 77 1 27 0 58 31 81 2 1 0 1 1 97 98 97 3 2 6 1 1 15 76 98 39 11 2 98
```

Block 2 (residues 46–90)

```
Secondary   46      50        C    60    4    70        80        D        90
Consensus   A L G I K M G E P W F K I K D L L R R H G V V F S S N Y A L Y A D M S N R V M S T L E E
Percent     52 70 87 64 48 75 80 25 94 41 86 50 29 48 26 37 22 23 33 37 51 63 25 49 59 94 89 94 85 34 94 100 46 70 57 96 22 86 60 73 23 64 46 28
AA type     A L h s l M h s P h a p . p t . h . t t . l l . s u n Y A L Y A D M S t + h h t h h t
E. coli     A L G V K M G P W F K Q K D L F R C G V V C F S S N Y E L Y A D M S N R V M S T L E E

A   214 39 1 0 55 9 30 67 5 23 0 4 39 15 42 19 33 14 30 21 4 8 69 97 4 21 0 139 1 0 188 7 1 2 65 6 3 18 65 8 13 58 17
C   0 0 0 0 0 0 0 0 48 0 0 0 1 1 105 16 1 68 1 1 23 0 0 46 15 0 1 0 1 0 0 0 0 0 0 1 1 0 1 0 0 0 0
D   10 0 3 0 0 0 0 0 0 0 0 1 0 0 1 1 3 0 8 12 3 22 0 0 0 0 0 5 0 3 1 0 0 288 0 0 15 1 0 0 16 0 0 36
E   32 0 0 0 11 0 0 101 0 9 0 0 28 1 28 82 27 5 90 30 23 9 0 0 15 0 1 0 0 118 0 0 5 31 0 0 25 4 0 0 45 4 0 190 115
F   3 19 2 0 0 0 0 0 0 35 355 0 0 0 0 0 0 21 66 13 0 3 0 3 9 244 0 0 0 0 38 0 3 2 0 6 0 0 53 11 0 0 6 1 0
G   0 2 359 0 9 4 329 10 0 0 0 0 0 0 1 0 10 0 0 2 3 7 211 5 0 0 0 1 5 0 0 170 1 0 3 11 0 3 4 0 0 55 18
H   0 1 5 0 3 0 0 0 0 0 23 3 1 2 31 2 0 5 3 152 11 0 33 0 0 0 0 2 3 0 1 0 0 2 3 0 18 3 0 1 3 9 0 3 4
I   0 1 15 262 0 0 0 8 0 17 0 1 118 4 0 41 66 0 0 99 72 30 3 0 1 0 0 2 3 0 3 0 43 0 1 0 22 9 0 84 79 0 1
K   34 0 10 1 196 38 1 13 0 1 2 205 2 197 23 26 60 55 34 42 0 22 7 0 1 0 0 2 3 0 0 0 5 0 3 4 0 9 29 0 1 0 21 17
L   16 289 0 33 1 0 0 3 5 34 1 14 41 6 0 153 89 32 3 0 34 10 12 63 0 0 0 385 5 0 99 5 1 27 2 3 59 265 8 5
M   1 10 0 0 0 310 0 19 0 4 0 5 5 1 0 19 7 0 3 2 0 0 34 0 2 0 0 20 0 4 0 0 236 0 3 7 7 299 1 5 2 0 4
N   6 0 6 0 5 2 1 1 0 0 0 3 0 22 6 0 3 0 13 3 32 64 0 5 0 3 0 387 0 8 0 2 15 0 6 90 0 13 30 1 6 9
P   21 0 1 0 70 12 0 1 386 0 0 0 0 5 0 0 0 0 1 0 10 16 0 1 0 21 0 0 6 0 0 5 0 4 0 0 0 0 0 0 2
Q   30 0 0 0 13 0 0 30 0 0 130 52 27 26 22 4 31 87 22 17 0 27 0 0 0 0 5 2 0 0 52 5 0 4 35 0 2 32
R   17 2 2 0 30 14 5 5 7 1 3 16 4 122 25 34 16 93 134 14 18 0 22 0 42 1 1 4 0 42 0 0 64 355 0 3 30 0 33 65
S   13 1 1 0 0 0 0 9 4 3 0 4 0 0 16 2 1 0 12 54 10 21 2 1 4 4 0 387 367 1 0 0 26 1 0 12 54 0 396 51 2 1 87 4 21
T   4 0 0 0 0 7 8 0 39 0 3 0 2 0 4 5 2 5 3 0 0 1 11 1 1 0 79 4 0 1 1 4 2 0 2 0 64 123 1 8 21
V   6 6 1 114 3 2 0 60 4 21 0 0 100 0 12 29 2 1 0 259 103 200 10 0 10 0 1 0 1 0 1 8 0 0 246 17 10 109 21 0 1
W   0 4 0 0 1 0 0 0 92 2 0 7 0 0 0 2 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 0
Y   6 12 3 1 1 1 0 0 168 1 0 9 0 0 2 4 7 4 8 0 1 8 0 22 0 0 0 348 0 3 409 0 3 9 0 0 2 0 0 21 1 0
X   8 17 0 1 1 0 0 0 0 0 0 0 1 3 2 2 2 1 2 1 1 0 0 8 0 0 0 0 0 0 9 0 0 3 0 0 4 7 0 0 0 1 1 1

HYDRO   262 395 8 410 131 336 74 158 400 394 379 24 344 33 66 269 371 81 64 100 15 404 268 404 361 11 38 2 407 164 401 411 202 11 404 2 74 12 409 363 91 270 408 68 30
LONG    64 96 2 100 32 82 18 38 97 96 92 6 84 8 16 65 90 20 16 25 4 98 65 98 88 3 9 0 99 40 98 100 49 3 98 0 18 3 99 88 22 66 99 17 7
SHORT   106 7 19 1 255 71 8 92 11 5 6 353 61 346 116 95 2 210 292 123 143 2 91 5 47 398 368 399 1 118 8 0 20 78 6 406 268 391 2 34 252 128 3 94 208
(lower) 26 2 5 0 62 17 2 22 2 1 1 86 15 84 28 23 51 71 30 35 0 22 1 11 97 90 97 0 29 2 0 5 19 1 99 65 95 0 8 61 31 1 23 51
```

Block 3 (residues ≈91–135)

```
Consensus   F A P R V E I Y S I D E A F L D L T G V P N C R D L T D F G R E I R A T V L Q W T G I P V
Percent     31 30 75 26 43 82 43 98 100 97 100 100 47 97 48 61 60 32 57 19 17 15 24 19 48 60 21 25 36 93 23 49 76 25 32 66 36 29 29 41 66 55 63 56
AA type     h s s t s r t Y S I D E A F h h p h t t . t h h . s t t h . h s t t h . h . p . h t h t V
E. coli     L S P R V E I Y S I D E A F C D L T G V R N C R D L T D F G R E I R A T V L Q R T H L T V

A   16 123 12 50 6 2 8 3 0 0 0 0 194 0 36 2 14 12 19 3 9 61 25 19 6 15 61 2 111 16 29 3 101 38 7 9 8 1 13 4 5 6
C   4 29 0 4 6 0 0 0 0 0 67 0 52 3 12 1 0 21 0 97 1 29 0 1 0 39 2 0 0 5 2 1 0 0 19 12 0 0 43
D   0 0 15 54 0 5 0 0 0 411 0 0 249 0 22 11 1 25 8 33 18 199 0 23 101 0 0 34 0 0 18 10 0 4 13 6 0 0 0
E   7 3 26 86 0 335 2 0 0 0 411 0 0 15 0 19 5 0 25 8 7 20 24 74 65 0 17 95 0 0 74 35 0 5 6 56 1 2 0
F   127 9 1 0 3 1 0 0 0 6 400 7 0 71 0 0 28 28 7 24 16 0 11 13 148 0 2 5 0 1 0 11 6 4 0 6 1 0
G   0 8 0 0 4 0 0 1 0 0 0 2 0 9 7 0 10 236 0 5 23 8 46 4 1 1 8 0 242 0 7 0 3 1 0 10 1 20 0 273 0 3 0
H   13 0 1 27 4 36 0 4 0 0 1 0 0 0 10 0 3 24 0 9 32 13 9 4 31 0 43 9 0 3 14 0 13 12 28 0 52 0 3 0
I   7 4 2 1 59 0 176 2 0 399 0 0 2 47 1 29 0 74 2 4 43 43 1 21 9 9 1 2 0 203 7 1 8 112 4 1 3 0 164 0 61
K   2 0 2 15 0 4 0 0 0 0 0 8 0 38 0 0 9 36 4 30 1 3 20 21 1 0 26 52 0 29 44 33 0 32 102 0 1 0 0
L   107 1 2 4 48 0 1 1 0 0 0 20 1 198 8 247 2 1 35 12 16 7 1 0 245 15 9 57 2 19 147 2 4 56 0 226 0 0
M   54 0 0 0 17 3 2 0 0 0 0 0 30 0 14 0 12 0 0 28 0 1 25 0 4 2 0 3 10 7 35 0 0 0 6 1 1 5 0 13
N   0 0 3 17 0 1 1 0 0 0 0 5 0 0 17 0 16 26 1 22 62 17 4 44 1 12 6 1 0 8 0 0 21 5 1 6 41 0 3 0
P   0 0 307 0 0 0 30 0 0 0 0 0 4 0 15 10 0 71 0 23 28 8 33 2 19 0 1 0 0 62 0 7 2 0 15 1 0 257 0
Q   10 0 3 15 62 0 38 0 0 0 0 0 5 0 0 0 25 0 44 35 1 55 14 17 80 1 14 23 14 0 26 0 15 62 87 33 40 30 12 121 19 0 0
R   4 0 1 106 0 0 14 0 0 0 0 0 0 0 25 0 44 35 1 55 14 17 80 1 14 23 14 0 175 26 0 311 57 77 71 113 58 0 45 0 3 0
S   6 96 27 6 12 8 1 1 410 0 0 0 0 75 0 40 84 17 13 11 38 18 13 58 0 29 15 0 2 26 5 1 25 1 74 17 1 51 0
T   2 41 5 8 10 2 1 3 0 0 0 3 0 0 24 0 0 132 6 7 17 7 7 11 9 3 87 23 6 4 11 19 2 11 130 2 1 4 3 250 16 0 112 44
V   4 77 2 6 178 1 135 0 0 4 10 0 0 9 0 7 45 1 25 2 1 77 7 24 16 2 25 4 32 7 1 2 270 2 2 39 0 16 6 231
W   2 14 0 2 1 1 2 0 0 0 0 0 0 0 0 0 0 0 2 0 46 0 119 0 0 0 0 0 0 0
Y   40 0 2 0 5 8 1 401 0 0 0 0 0 0 13 0 65 0 1 0 26 2 5 7 11 0 2 5 0 1 0 1 11 0 0
X   2 0 2 1 0 5 0 0 0 0 0 0 0 1 5 0 80 58 12 0 0 1 80 9 10 18 36 8 7 3 12 6 1 2 16 31 1 23 51

HYDRO   361 257 328 68 323 24 350 408 7 4 1 4 411 0 326 411 401 25 411 33 31 292 140 110 262 167 48 358 86 115 345 158 57 66 409 30 109 71 409 258 12 158 159 4 411 277 360
LONG    24 6 9 153 62 1 57 0 0 0 0 0 79 0 5 0 57 0 105 69 3 99 127 53 125 50 18 81 67 27 0 271 171 0 373 162 145 0 116 342 125 0 51 0 5
SHORT   30 143 41 167 84 11 59 3 410 0 0 0 83 0 1 105 0 321 92 8 127 172 78 149 117 21 188 115 33 11 291 205 2 375 199 280 2 121 367 133 251 80 1 128 51
(lower) 7 35 10 41 20 3 14 0 100 0 0 0 20 0 0 26 0 78 22 2 31 42 19 36 28 5 46 28 8 2 71 50 0 91 48 68 0 29 89 32 61 19 0 31 12
```

Block 4 (residues 136–180)

```
Secondary   136  7  140      F      150        160    8        170      G        180
Consensus   G V G I A P T K T L A K L A N H A A K S . K P Q T G V V D L S N P E R R R K L L A L L P
Percent     37 67 99 73 51 43 80 83 84 87 92 100 54 90 81 46 83 98 52 29 25 24 33 36 87 82 34 43 63 23 28 24 20 34 30 24 30 50 73 24 26 23 45
AA type     s G h u . o h s h h A K h u s h h A A h . K p . . h s u s h h t s N p . h h h L M s h . h
E. coli     G V G I A Q T K T L A K L A N H A A K S . K P W Q T G V V D L S N P E R K L M S A L P

A   17 5 4 4 209 48 1 2 11 0 380 0 26 369 1 1 188 343 2 7 1 7 36 72 10 40 2 0 5 112 5 2 19 55 5 15 9 17 87 39 3 1
C   105 0 0 0 0 0 0 0 0 12 15 1 4 0 2 0 0 5 0 112 0 2 10 1 0 0 0 0 19 0 1 0 0
D   0 0 0 0 0 0 0 23 0 0 0 0 0 0 2 12 19 0 30 16 1 0 175 0 77 107 33 53 31 2 60 19 0 34 0 2 36
E   0 2 0 0 0 0 23 0 0 4 0 15 1 7 4 0 17 15 23 26 0 0 1 1 30 0 0 74 4 84 34 9 68 19 0 34 8 2 30
F   0 2 0 0 0 0 0 15 0 0 2 15 0 1 57 2 19 11 0 1 0 1 0 12 0
G   154 0 407 0 186 2 0 0 0 0 0 0 0 0 2 5 0 73 1 146 359 0 21 11 12 5 0 5 2 2 0 43 0
H   0 0 0 0 0 0 0 220 0 0 0 5 16 3 15 3 38 16 11 0 33 0 20 84 14 0 32 15 2 41 15 3 17 2 0
I   0 122 0 300 0 0 0 5 0 60 0 2 5 5 60 11 0 0 38 16 11 33 0 0 20 84 0 41 26 4 42 13 0 84 89 0
K   0 0 0 0 12 0 0 0 0 411 0 0 0 33 2 404 214 0 10 31 20 0 0 22 42 9 5 31 124 1 53 26 44 0
L   0 0 0 29 0 1 0 40 6 357 0 222 1 0 0 0 0 5 68 0 0 39 15 258 3 29 12 16 45 33 5 206 298 5 105 93 5
M   0 0 0 21 0 0 0 0 0 30 0 0 14 0 0 0 0 0 0 0 7 4 76 3 8 0 15 0 7
N   0 0 0 0 0 0 2 0 29 0 332 0 4 0 0 0 0 2 1 4 39 0 2 43 0 12 117 2 18 38 0 13 0 2 0
P   0 0 0 0 178 0 13 0 0 0 0 0 5 0 0 0 0 66 101 0 14 0 0 98 16 1 19 0 5 15 5 2 186
Q   0 0 0 0 58 0 1 26 0 31 3 1 33 0 33 0 41 1 2 0 66 0 26 0 89 20 16 0 35 57 32 1
R   0 0 0 0 0 21 0 0 1 0 0 0 52 1 0 38 1 12 54 20 41 15 54 140 122 99 0 0 1 6 0 5
S   99 0 0 0 1 16 51 1 9 0 0 28 0 28 0 33 0 0 17 17 18 0 96 39 8 23 0 1 74 17 1 50
T   21 0 0 0 1 0 329 3 246 2 0 12 15 14 0 2 136 0 0 339 141 63 10 18 29 16 75 39 2 30 76 2
V   0 275 0 0 26 0 0 0 0 0 55 0 0 2 25 2 0 0 5 0 0 58 0 7 1 0 18 6 0 5 5 0 54 5
W   0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 57 0 0 4 0 0 1 0
Y   0 12 0 0 0 0 0 0 0 0 0 0 0 0 26 0 0 0 0 0 0 18 62 0 7 36 15 5 6 15 5 1 1 0
X   0 0 0 0 0 0 0 0 0 0 0 18 228 36 95 15 12 9 14 1 0 1 23 15 7 1 6

HYDRO   122 411 4 410 209 241 2 61 40 368 382 0 383 396 1 22 172 192 152 103 189 60 3 395 387 163 409 72 33 205 121 72 147 82 107 402 410 199 197 289 249 61
LONG    15 0 0 0 0 23 7 85 0 9 3 0 35 0 41 0 408 298 2 113 93 151 20 105 20 5 3 1 0 2 0 129 172 14 0 4
SHORT   135 0 0 1 16 131 409 350 356 35 29 411 28 5 393 92 15 52 508 340 1 139 119 188 172 137 25 12 22 60 1 211 222 70 121 263 247 189 241 7 1 206 184 119 82
(lower) 33 0 0 0 4 32 100 85 87 9 7 100 7 1 96 22 4 13 0 83 0 34 29 46 42 33 6 3 5 15 0 51 54 17 29 64 60 46 59 2 0 50 45 29 20
```

FIGURE 2: Continued.

Figure 2: Continued.

Block 1 (positions 361–405)

	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405
Secondary	361									370										380										390										400					405
Consensus	F	D	D	N	A	P	R	P	K	S	E	A/K	L	M	A	V	L	D	K	L	N	A	R	Y	G	R	G	T	L	R	F	A	G	Q	G	I	Q	Q	Q	W	Q	M	R	R	E
Percent	80	30	24	12	20	34	33	21	23	41	31	19	88	83	18	38	39	90	12	43	85	21	45	27	90	38	53	46	44	15	34	69	31	36	71	29	14	26	15	82	27	75	41	58	38
AA type	h	D	D	N	A	P				t	t				t	l			t	h	h		p		u							u								h		t		p	t
E. coli	F	D	D	N	A	D	N	A	P	Q	L	M	T	V	M	D	T	N	A	K	E	G	R	G	u	R	G	T	L	R	F	A	G	Q	G	I	Q	Q	Q	W	Q	M	R	R	A
A	1	12	16	43	81	19	10	64	13	17	59	80	6	2	76	92	2	0	48	2	0	85	2	2	0	10	7	45	3	7	12	285	103	33	8	8	39	15	21	4	112	4	2	20	52
C	0	7	2	2	0	2	2	0	2	2	6	0	1	5	0	1	3	6	0	0	0	17	1	0	3	0	0	0	0	0	0	4	25	1	1	1	1	0	1	0	1	7	6	0	52
D	0	124	100	26	22	46	25	15	18	17	35	6	1	9	7	0	1	369	12	0	1	17	1	18	0	3	49	0	0	0	0	1	3	9	4	43	6	38	3	6	0	0	1	0	52
E	0	16	59	21	71	44	13	67	13	10	126	20	0	5	42	0	0	10	30	0	0	11	3	34	0	2	11	0	0	1	2	57	4	1	31	8	50	0	12	1			2		158
F	328	4	4	4	2	0	5	0	0	1	0	0	0	3	0	0	3	0	0	0	4	0	3	0	84	0	4	2	11	0	2	0	60	138	1	1	1	2	13	0	0	1	20	1	4
G	0	11	10	13	2	9	6	68	10	9	5	0	4	14	0	0	0	14	0	0	43	0	13	1	4	369	2	217	1	1	27	0	86	126	2	290	0	27	6	24	0	29	2	1	16
H	0	2	6	7	2	10	5	22	15	14	8	5	0	0	11	0	0	1	33	0	6	14	18	16	3	9	13	2	0	0	12	5	0	2	6	0	6	12	9	4	3	18	0	5	3
I	9	6	0	16	5	6	12	10	0	6	6	1	7	1	0	11	76	0	3	176	0	0	2	0	0	87	4	19	0	0	1	4	0	0	0	1	0	10	0	1	0	11	3	0	2
K	0	5	1	2	14	16	27	31	93	19	30	80	3	0	57	1	1	0	49	0	4	25	132	2	8	109	17	65	4	23	4	0	1	4	10	5	27	60	49	0	37	0	162	10	18
L	41	5	3	14	9	16	25	6	0	7	4	0	363	7	10	17	162	2	8	140	0	8	4	6	0	13	0	6	180	10	99	6	11	7	9	49	3	0	2	3	23	2	0	4	
M	3	4	1	1	2	1	4	2	0	1	3	0	1	343	2	86	0	2	6	0	0	19	0	4	0	2	11	4	8	0	0	5	3	33	4	2	0	4	310	0	0	0	0	0	
N	1	8	9	51	30	4	14	9	48	28	9	14	0	1	17	0	0	2	14	0	351	15	6	8	7	20	31	4	0	3	0	0	4	2	5	3	33	4	2	0	3	17	2	28	
P	0	16	24	23	26	138	23	86	0	1	35	10	0	0	0	0	0	2	39	0	4	49	31	1	1	6	3	11	0	22	2	0	2	150	5	3	59	105	63	0	83	3	8	81	34
Q	0	12	10	19	34	44	25	18	24	15	32	69	0	10	63	1	0	2	39	0	49	31	1	1	3	11	0	22	0	2	150	5	3	59	105	63	0	83	3	8	81	34			
R	0	2	1	15	9	4	135	12	79	56	16	65	1	0	32	0	0	1	38	0	22	37	185	10	0	156	4	8	0	62	8	1	10	27	3	7	20	42	7	3	18	0	169	237	3
S	0	7	49	27	26	9	31	18	10	170	9	19	0	8	36	11	0	1	22	1	3	17	3	47	6	28	32	38	0	14	14	8	71	24	12	8	43	10	40	0	35	3	13	16	11
T	0	13	14	16	31	14	21	16	5	17	9	12	0	3	18	98	4	0	49	8	3	21	0	17	2	189	0	20	11	1	7	21	1	34	25	6	19	4	18	9	4	0	1	2	
V	0	3	1	38	16	13	4	10	1	1	2	5	11	0	8	156	28	0	49	0	0	4	2	0	4	0	6	109	21	23	4	18	18	1	24	11	0	5	0	2	4	2	1	2	
W	12	0	0	0	0	0	5	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	337	0	1	0	0	0
Y	1	3	3	8	6	3	0	2	1	1	0	0	4	0	0	0	3	0	1	0	0	2	3	0	5	0	51	11	0	6	0	6	0	12	0	6	24	0	29	2	1	1	0	1	0
X	15	151	98	65	23	16	16	17	22	17	17	17	15	15	16	16	16	16	91	17	19	17	17	21	16	17	17	17	19	21	75	30	32	25	25	24	22	19	19	19					

Block 2 (positions 406–422)

	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422
Secondary	406				410										420		422
Consensus	M	L	S	P	R	Y	T	T	R	W	S	D	L	P	V	V	K
Percent	31	62	78	69	28	73	83	84	50	49	23	45	52	55	23	45	35
AA type		h	s		t				s	t		t	h			h	
E. coli	M	L	S	P	R	Y	T	T	R	S	S	D	L	L	V	V	K
A	1	1	0	5	84	18	2	13	7	3	37	1	0	9	16	97	4
C	0	1	1	0	22	2	0	9	0	3	10	2	0	3	1	1	
D	2	0	6	0	9	7	0	1	37	0	63	185	0	2	3	0	4
E	4	0	1	2	5	0	3	1	3	2	28	110	1	1	26	0	6
F	21	3	0	1	1	15	2	4	0	21	1	1	0	0	1	2	6
G	1	0	7	14	19	1	0	4	0	1	16	7	0	3	5	1	11
H	45	3	0	1	13	9	0	1	5	0	6	0	0	0	1	3	15
I	1	2	6	0	9	0	0	0	31	0	112	29	28	69	9		
K	18	31	2	21	8	2	4	0	3	22	2	0	2	34	1	142	
L	56	254	0	2	1	0	15	0	3	51	4	1	214	82	17	9	9
M	128	14	0	0	0	1	0	0	4	2	0	15	4	3	0	5	
N	19	0	2	6	21	5	4	0	21	2	37	7	0	0	3	0	14
P	2	0	0	283	0	5	6	1	0	0	0	1	224	1	0	1	
Q	22	3	1	23	39	0	0	0	24	0	17	36	0	0	12	0	16
R	36	71	1	23	117	2	0	0	204	1	27	0	0	0	74	0	85
S	9	7	322	6	40	10	6	14	31	20	94	14	0	1	6	2	14
T	4	1	44	1	3	1	341	345	1	0	12	2	0	0	44	2	11
V	0	5	3	0	0	0	0	2	8	1	0	31	5	96	186	12	
W	6	0	0	0	0	0	0	0	202	1	0	0	13	0	0	2	
Y	20	0	0	0	0	301	1	1	0	31	5	0	0	0	3		
X	16	15	15	23	20	25	24	26	33	32	32	34	36	36	38	38	41

FIGURE 2: Preferred amino acid at the 408 amino acid positions in UmuC(V), which is the polymerase subunit of DNAP V. Top row: amino acid numbering (every tenth residue) and the likely secondary structure based on alignment to other Y-Family DNAPs (β-strand, yellow; α-helix, blue; turn, green). Second row: consensus amino acid. Third row: percentage of the consensus amino acid. Fourth row: consensus type of amino acid, as codified in [101] (Website: http://coot.embl.de/Alignment/consensus.html). Regions where the consensus is high are highlighted in pink (all Y-Family DNAPs) or red (UmuC(V) only). Fifth Row: UmuC(V) sequence from *E. coli*. Rows 6–25 show the number of amino acids of each type at each position using the conventional one-letter code. Row 26: We note that some UmuC(V) sequences have slightly more or fewer than 422 amino acids. The number of UmuCs with an amino acid missing at a position compared to *E. coli* is shown in Row 26 ("X"). When an amino acid is present in a UmuC(V) but absent in *E. coli*, then the sequence is merely not included in Figure 2 (see text). In terms of gaps, *E. coli* only has one position that does not conform to the majority of UmuC(V)s: *E. coli* has an amino acid at position 156 (W), which is not present in a slight majority of all UmuC(V)s (226/408). The 408 UmuC(V) sequences were taken (as of 6/21/10) from the databases UniProt/Trembl [102] and HAMAP [103]. The sequences were aligned using MUSCLE [104]. Sequences were excluded for three reasons: (1) they were incomplete (i.e., sequences without recognizable versions of the N-terminus and the C-terminus were excluded); (2) they lacked any of the catalytic residues D6, D101, or E102; (3) they were redundant within a species (i.e., within a species if two UmuCs had the identical sequence, then only one was included).

C8-dG, as do most aromatic amine mutagens/carcinogens, where AAF and AAF-C8-dG have frequently been used as models to probe the mutagenic and carcinogenic mechanisms of aromatic amines [117]. In *E. coli* AAF has a major mutational hot spot in 5′-CG_1CG_2 sequences in which it induces −2 frameshift mutations [117]. AAF-C8-dG at G_2 (but not G_1) causes a −2 frameshift mutation in a DNAP II-dependent process, or causes no mutation in a DNAP V-dependent process [117]. The current model is that AAF-C8-dG at a replication fork exists in two different conformations [66, 117]. In one conformation, the adducted dG moiety is in a −2 slipped intermediate, which DNAP II uses for insertion, and then in the presence of β-clamp an additional three extension steps (to L + 3) are accomplished, at which point replication can be successfully continued by DNAP III [66]. From a nonslipped intermediate, DNAP V inserts dCTP opposite AAF-C8-dG and then extends by adding four more dNTPs (to L + 4), after which DNAP III can successfully continue replication [66]. These two pathways are followed approximately equally in cells, though by manipulating the concentration of DNAP II versus DNAP IV, the ratio [−2 frameshift : no mutation] can be modulated, suggesting that the two conformations interconvert [93]. In vitro in 5′-CG_1CG_2 sequences DNAP II also does TLS to give a bypass product that should ultimately yield a −1 frameshift mutation, which are not, however, observed in vivo; recent in vitro studies suggest that DNAP II cannot extend far enough from the −1 frameshift intermediate, and, thus, the 3′ → 5′ exonuclease activity of DNAP III degrades the intermediates in the −1 frameshift pathway [66].

Molecular modeling has provided insights about how lesion bypass might occur; for example, a modest alteration in sugar pucker (from C3′-endo to C1′-exo) is required before AAF-C8-dG can Watson-Crick base pair with dCTP [94, 119]. Though this work was done in Dpo4 and hDNAP ι, there is every reason to think that a similar conclusion would be reached for DNAP V. Recently, X-ray structures of the corresponding deacetylated adduct AF-C8-dG has

been reported [32]. Regrettably, the structures do not reveal insights about how dCTP might be inserted opposite AF-C8-dG, but they do offer a glimpse of more-or-less normal Watson-Crick AF-C8-dG : dC base pairing in the L + 1, which has the AF moiety in the opening on the major groove side of Dpo4, and in the L + 2 position, in which the AF-moiety is accommodated by a modest rearrangement in the little finger domain.

Benzo[a]pyrene (B[a]P) is a well-studied DNA damaging agent that is a potent mutagen/carcinogen and an example of a polycyclic aromatic hydrocarbon (PAH), a class of ubiquitous environmental substances produced by incomplete combustion [120, 121]. PAHs in general and B[a]P in particular induce the kinds of mutations thought to be relevant to carcinogenesis and may be important in human cancer [122–128]. B[a]P mutational spectra were established with the major metabolite that reacts with DNA (i.e., (+)-*anti*-B[a]PDE), in *E. coli* [129], yeast [130, 131] and mammalian (CHO) cells [132]. Mutagenesis has also been studied with [+ta]-B[a]P-N^2-dG (+BP, Figure 1), the major adduct of (+)-*anti*-B[a]PDE, and G → T mutations predominate in most cases (see [133] and references therein).

DNAPs IV and V of *E. coli* are both involved in TLS with B[a]P-N$_2$-dG adducts, although they play very different roles. In studies with purified proteins, DNAP IV inserted dCTP (>99%) opposite both +BP and its mirror image −BP ([-ta]-B[a]P-N^2-dG) in a 5′-C**G**A sequence, while DNAP V inserted dATP (>99%) [77]. This tendency is evident in *E. coli*. DNAP IV is required in the nonmutagenic pathway with +BP [72–75], −BP [75] and other N^2-dG adducts [72, 76]. An amino acid change (F12I) at the conserved "steric gate" (which excludes rNTPs) decreases dCTP insertion in vitro opposite several N^2-dG adducts and similarly decreases TLS in vivo, which argues that DNAP IV does dCTP insertion in vivo [72]. In the nonmutagenic pathway DNAP V is required in addition to DNAP IV with +BP [73–75]. Why are two DNAPs required for nonmutagenic TLS with +BP: certain lesions need one DNAP for insertion and a second for extension [134, 135]. Thus, if DNAP IV does dCTP insertion [72–77], then DNAP V must do extension, which is sensible given kinetic findings with purified proteins show that DNAP V can be significantly better than DNAP IV at the step directly following adduct-G : C formation (i.e., extension) in the case of +BP compared to −BP (discussed in greater detail in reference [75]). Regarding the nonmutagenic pathway with −BP, only DNAP IV is required for efficient TLS [75], suggesting it does both insertion and extension. In a 5′-T**G**T sequence, DNAP V is required in the G → T pathway for +BP, while DNAPs II and IV are not, implying that DNAP V must do insertion and extension [57]. However, in a 5′-G**G**A sequence, G → T mutations were shown not dependent on DNAP V and were not enhanced by SOS induction, which implies no lesion-bypass DNAP involvement and led the authors to propose that DNAP III was involved in dATP insertion opposite +BP [73, 74]. Random mutagenesis studies with [+*anti*]-B[a]PDE also showed the existence of a non-SOS-inducible G → T pathway (discussed in [57]), though the major G → T pathway did require SOS-induction, implying involvement of a lesion-bypass DNAP.

7. Architecture of Y-Family DNAPs

Table 1 [44] shows that dNTP insertion opposite a variety of adducts/lesions, including +BP, is remarkably similar for the DNAP IV and DNAP κ pair, suggesting they are functional orthologs. Insertion is also remarkably similar for the DNAP V and DNAP η pair, suggesting they are also functional orthologs. There must be structural reasons for the insertion preferences of these DNAPs, though the key elements are not obvious, given that in alignments, for example, UmuC(V) shares only 20% amino acid identity with its functional ortholog hDNAP η, which is about the same as the 21% identity that it shares with its nonfunctional ortholog hDNAP κ [44]. The extent of this dilemma is further revealed by the fact that hDNAP η is no more identical to scDNAP η (24%) than it is to hDNAP κ (24%). Nevertheless, a careful examination of Y-Family DNAP structure suggests that key structural features do exist.

A variety of architectural features are revealed by considering how B[a]P-N^2-dG adducts must sit in the active sites of Y-family DNAPs and how these structures might relate to adduct processing [136]. To form an adduct-dG : dCTP base pair, the B[a]P moiety must be in the developing minor groove, since the adduction site (N^2-dG) is in the minor groove in a Watson-Crick base pair. On the minor groove side, Y-Family DNAPs have an opening (or gap) next to the active site between the fingers and little finger domains. This opening looks like an elliptical hole of varying sizes in Dpo4 [24–32], Dbh [23], hDNAP ι [36–38] and in models of DNAP IV and UmuC(V) [44], while it looks like a slot in hDNAP κ [35]. It is not unreasonable to think that the size and shape of this opening might influence dNTP insertional mechanism given that the bulky B[a]P moiety must interact with this opening on the minor groove side.

The character of this opening can be analyzed based on a simple analogy to a "chimney." Three regions of the protein contribute to the chimney as shown in Figure 3(a) for our model of DNAP IV: an upper lip (aa33–36, turquoise) and a left lip (aa73–76, blue), which are in the fingers domain, while the lower lip (blue, aa244–247), is in the little finger domain [136]. The UmuC(V) chimney is shown in Figure 3(b).

Two features control the size and character of this opening. (1) The amino acid side chains in the upper lip (aa33–36 in DNAP IV) can be thought of as a "flue," which either plug the chimney leaving a small opening or do not plug the chimney leaving a large opening. The flue amino acids are present in all Y-Family DNAPs. (2) A "cap" may lie over the top of the chimney opening. The cap is formed by an insert of amino acids in the left lip of the chimney and is only preset in DNAP η [39, 40].

First, we consider how the "flue" amino acid side chains influence the character of the "chimney." Why do we think that the chimney is a key structural feature that is likely to be important for protein function? If the chimney is important, then evidence for its importance should exist, even in the case of UmuC where no X-ray structure exists. We aligned 408 UmuC(V) sequences, and Figure 2 shows the total number of each of the twenty amino acids that are

(a)

(b)

(c)

(d)

(e)
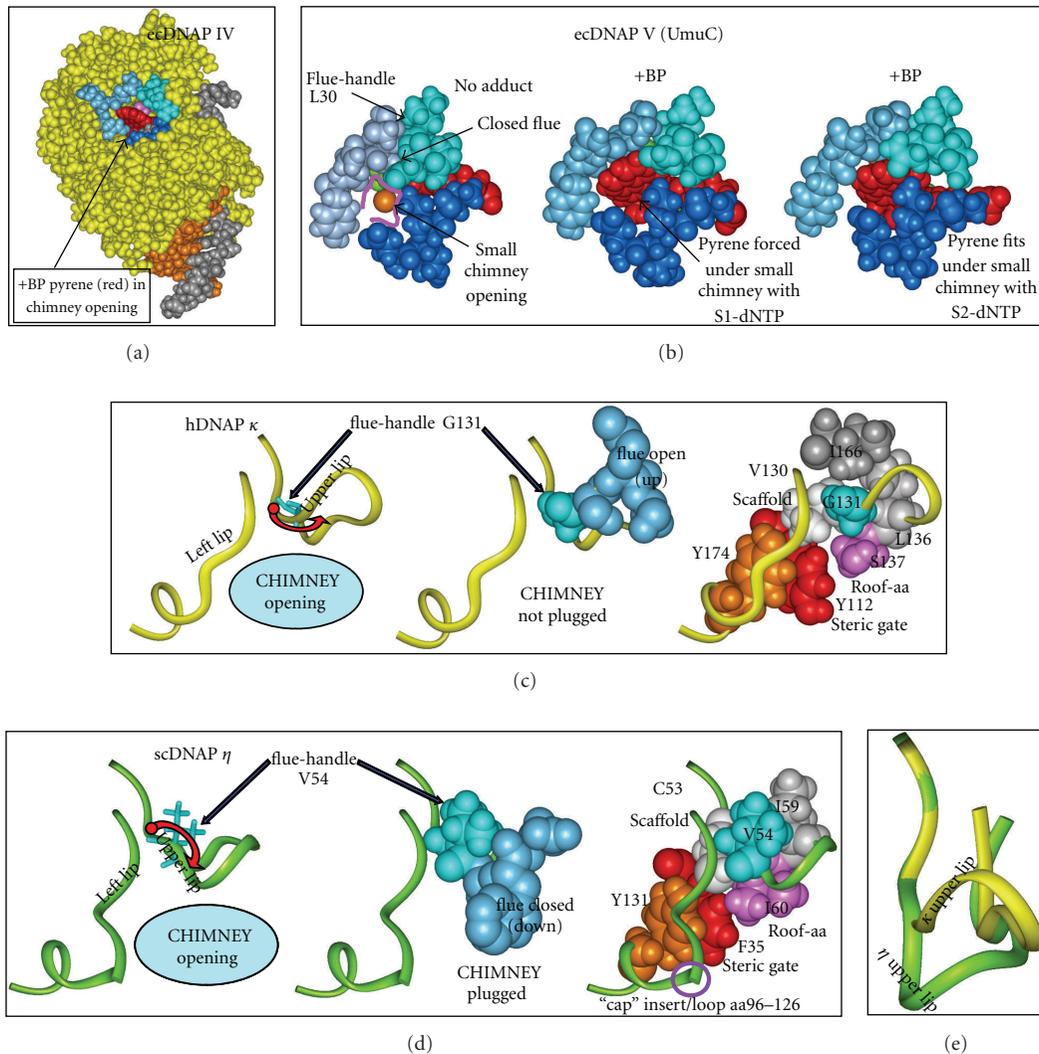
FIGURE 3: Structures of regions of ecDNAP IV (a), ecUmuC(V) (b), hDNAP κ (c), scDNAP η (d). (a) View from the minor groove side of DNAP IV (yellow), showing the "chimney" opening (cleft or hole), which is defined by the upper lip (turquoise, aa33–36), left lip (blue, aa73–76) and lower lip (dark blue, aa244–247). The chimney opening in DNAP IV is large enough to accommodate the pyrene moiety of +BP (red). In DNAP IV, the dG moiety of +BP can base pair comfortably with dCTP in the canonical S1-dNTP shape. (Neither the dG moiety of +BP nor the dCTP are visible.) The template (gray) and primer (brown) are also shown. (b) Models of UmuC(V) with no adduct (B/left) or +BP paired with dCTP in the canonical "chair-like" S1-dNTP shape (b/center) or +BP paired with dCTP in the non-canonical, "goat-tail-like" S2-dNTP shape (b/right). The chimney opening is small (b/left) and the pyrene moiety of +BP does not fit into the chimney such that the S1-dNTP shape is possible (b/center). In contrast, the pyrene moiety of +BP fits under the chimney opening in the case of pairing with dCTP in the S2-dCTP shape (b/right), because it sits lower down in the active site compared to S1-dNTP. (c) Regions of hDNAP κ. Y-Family DNAPs in the IV/κ-class have a glycine "flue-handle," such as G131 (turquoise, c/left) in hDNAP κ G131 adopts ϕ/φ-angles that lead to upward curvature of the protein backbone in the chimney upper lip (red arrow, c/left) and results in the "flue" amino acids (S132/M133, blue, c/center) pointing away from the chimney, giving a large opening. The c/right structure shows V130 (white) that serves as a scaffold to organize the chimney's upper lip and left lip (yellow ribbons), along with the roof-aa (S137, purple), the steric gate (Y112, red) and a conserved tyrosine (Y174, brown), which stacks on the backbone of the left lip and helps orient it. V130 forms a square with the G131 flue-handle, L136 (gray) and the S137 roof (pink) upon which I166 stacks (dark gray). (d) Regions in scDNAP η. Y-Family DNAPs in the V/η class have a bulky "flue-handle," such as V54 in scDNAP η (turquoise, d/left), which causes downward curvature of the chimney upper lip (red arrow, d/center), and results in the "flue" amino acids (Q33/W34, blue, d/center) plugging the chimney, giving a small opening. The structure in d/right shows scaffold C53 (white) organizing the chimney's upper lip and left lip (green ribbons in (c)), along with the roof-aa (I60, purple), the steric gate (F35, red) and a conserved tyrosine (Y131, brown), which stacks on the backbone of the left lip and orients it. C53 forms a square with the V54 flue-handle (turquoise), I59 (gray) and the I60 roof (pink). scDNAP η has a large insert/loop (aa96–126) in the left lip, which is represented as a discontinuity. (e) The upper lip for hDNAP κ (yellow) is superimposed on the upper lip for scDNAP η (green), which clearly shows the differences in curvature. X-ray coordinates are from 2OH2 for hDNAP κ [29] and from 1JIH for scDNAP η [27], where hydrogens were added using insightII.

found at each aa position. Positions with ≥90% aa homology or with clusters of high homology are highlighted in pink and red. Many of the first ~20 aa show high homology, including the presumptive catalytically essential aspartate (D6), and the steric gate (Y11). The region around the catalytically essential asparate/glutamate pair (D101/E102) is also highly conserved. These and other regions that are conserved in all Y-family DNAPs are highlighted in pink in Figure 2. Regions conserved in UmuC(V), but not in other Y-Family DNAPs, are highlighted in red in Figure 2. One such conserved region is V29-C36, which is part of a loop that includes one edge of the chimney lip (S31-D34). (This loop is discussed at greater length in Section 7.1). A second conserved region is S71-Y77, which includes the second lip of the chimney, as well as other features discussed below. This conservation is strong evidence that the nature of the chimney opening is important. The third edge of the chimney (E255-T258) is in the little finger domain, and in our model of UmuC the third lip is farther from the active site and appears less likely to impinge on adducts protruding from the minor groove. Consistent with this view the third lip of the chimney is less well conserved. Preliminary analysis of the chimney lips of large collections of DNAPs IV, κ, and η sequences also reveal considerable amino acid conservation of the chimney lips.

7.1. Structural Basis for a Large versus a Small Chimney Opening.
DNAP IV has a large chimney opening (Figure 3(a)), which can accommodate the pyrene thus allowing +BP to readily pair with dCTP when dCTP adopts the canonical shape observed in all other families of DNAPs [136, 137]. In contrast, UmuC(V) has a small chimney opening (Figure 3(b)), which forces +BP downward in the active site into a position where catalysis seems unlikely to be facile [136, 137]. What structural difference(s) in DNAP IV versus UmuC(V) might result in a large versus a small chimney opening, and is this structural difference(s) conserved in other Y-Family DNAPs in the IV/κ-class versus the DNAP V/η-class?

The chimney upper lip (turquoise, Figure 3(a)) is closest to the active site, and principally defines whether the chimney can accommodate the bulky B[a]P moiety. The first amino acid in the upper lip of *E. coli* DNAP IV is glycine (G32). We have collected 434 DNAP IV sequences from the literature, and 418 have glycine at this position. Furthermore, 13/13 DNAP κ proteins from different species have glycine at this position. The one X-ray structure for the IV/κ-class is hDNAP κ [29], which shows that this glycine (G131, turquoise, Figure 3(c)), is followed by upward curvature of the chimney upper lip (red arrow, Figure 3(c)/left). This glycine can be thought of as a "flue-handle" whose φ/φ-angles permit this upward curvature (see below), with the consequence being that the R-groups on the next several amino acids (the "flue"; S132/R133, blue in Figure 3(c)/middle) point away from the chimney opening, which remains open. Our models of DNAP IV also have this upward curvature (Figure 3(a)) with an open flue, which depends on the analogous glycine flue-handle (G32).

In contrast, leucine (L30 in Figure 2) is the flue-handle in UmuC(V) in 370/408 cases. Furthermore, 11/11 DNAP

η proteins from different species have a bulky valine at the flue-handle position. The X-ray structure of scDNAP η [25], which is in the in the V/η-class, shows that its bulky V54 flue-handle (turquoise, Figure 3(d)/left) is associated with downward curvature of the chimney upper lip (red arrow), which forces the "flue" (Q55/Y56, blue in Figure 3(d)/middle to plug the chimney. Figure 3(e) shows the upward curvature of the upper lip of hDNAP κ (yellow) superimposed on the downward curvature for scDNAP η (green). In UmuC(V) the sequence is slightly different (VLSN), though the outcome is the same: the bulky L30 flue-handle causes downward curvature, and an asparagine (N32) plugs the chimney giving a closed flue (Figure 3(b)).

Upward versus downward curvature of the chimney upper lip can be traced to the φ/φ-angles adopted by the flue-handle [136]. The φ/φ-angles for the nonglycine flue-handles in scDNAP η, hDNAP η, hDNAP ι, UmuC(V), and Dpo4 are all similar, resulting in downward curvature of the chimney's upper lip, causing the flue to plug the chimney and the chimney opening to be small. In contrast, Glycine has greater flexibility in its φ/φ-angles compared to all other amino acids, and the glycine flue-handles in hDNAP κ and DNAP IV adopt φ/φ-angles unique to glycine that allow upward curvature of the chimney's upper lip, which keeps the nearby flue amino acids away from the chimney opening.

7.2. Roof-aa and Roof-Neighbor-aa.
Another key difference between the IV/κ-class and the V/η-class is the bulk of the roof-aa (pink in Figures 3(c)/right and 3(d)/right), which is a positionally conserved residue that lies above the nucleobase of the dNTP, as seen in the active site of Dpo4 [24–32], yDNAP η [34], hDNAP ι [36–38], hDNAP κ [35], and hDNAP η [39, 40]. Isoleucine is the dominant roof-aa in UmuC(V) (227/408), with valine (156/408) being the next most prevalent aa (Figure 2). In fact, compared to *E. coli* wt-UmuC (100%), the mutant I38V-UmuC (137%) is slightly more active in the nonmutagenic pathway with +BP, while amino acids that do not branch at the β-carbon, including leucine, show much lower activity [138]. Immediately after the roof-aa in UmuC(V), principally alanine is found (346/408). In the case of DNAP η from different species, the [roof-aa/next-aa] is [I/A] in 10/11 cases. In the yDNAP η X-ray structure [I60/A61] form a hydrophobic layer above the nucleobase of the dNTP.

In the collection of 434 DNAP IV sequences, there is more variability at the equivalent [roof-aa/next-aa] positions: [S/T] is preferred (238/434, 59%), though any of the nonbulky amino acids S, A, or T can be found at both the roof-aa (434/434) and the next-aa (406/434). For DNAP κ, the roof position is also principally S, A, or T (10/13) and the next amino acid is always threonine (13/13). In X-ray structures the threonine methyl group in Dpo4 (T45) and in hDNAP κ (T138) sit near the roof-aa (A44 and S137, resp.), and the hydroxyl of the threonine forms a hydrogen bond with a nonbonded oxygen on Pβ of the dNTP.

When the [roof/next-aa] were mutated in wt-UmuC(V) from [I38/A39] to the single mutants [I38A/A39] or [I38/A39T], polymerase activity declined significantly; however, the double mutant [I38A/A39T], is nearly as active

as wild type UmuC(V) [138]. I38A/A39T-UmuC has the same sequence as wt-Dpo4 (A44/T45), which is in the IV/κ-class. These findings show the importance to activity of the coupling of the identity of the [roof-aa/next-aa].

7.3. The Interconnected Architecture of the Chimney and Roof Regions. To understand the interconnected architecture of the chimney/roof regions of Y-Family DNAPs, it is useful to focus on a bulky aliphatic amino acid, which is highly conserved V29 (374/408) in our collection of UmuC(V) sequences. (In the equivalent position, all 434 DNAP IV sequences have either valine or isoleucine; valine is present in 10/11 DNAP η sequences and in 13/13 DNAP κ sequences.) This amino acid plays a scaffolding role as revealed in X-ray structures [23–38] and in models [44, 136, 138]. Using hDNAP κ [35] as an example, this scaffolding valine (V130, white in Figure 3(c)/right) is the beginning of a loop that ends with the roof-aa, and the two form a backbone hydrogen bond (scaffold-C=O : HN-roof). This backbone hydrogen bond is also observed in X-ray structures from Dpo4 [24–32], scDNAP η [35], hDNAP κ [35], and hDNAP ι [36–38]. The scaffold-V130 also contacts the flue-handle (G131) and L136 (Figure 3(c)/right, gray). Thus, the base of this loop is anchored by a square of four amino acids (V130/G131/L136/S137). In scDNAP η, this region looks similar with C53/V54/I59/I60 (Figure 3(d)/right). The square includes I31/G32/I41/S42 in DNAP IV, and V29/L30/V37/I38 in UmuC(V). Evidence suggests that V29 and I38 are likely to be in contact in UmuC(V) [138].

Scaffold-V130 in hDNAP κ (white, Figure 3(c)/right) also helps organize the steric gate (Y12, red), which face-stacks with Y174 (brown), a highly conserved tyrosine whose other aromatic face contacts the backbone of the left lip of the chimney (i.e., aa168–171 in hDNAP κ), thus helping to orient it. A tyrosine is found at this position in 406/408 UmuC(V) sequences, in 432/434 DNAP IV sequences, in 11/11 DNAP η sequences, and in 13/13 DNAP κ sequences.

7.4. The Chimney "Cap". The interconnection between the roof and chimney regions are similar in scDNAP η (Figure 3(d)/right). However, the left chimney lip has an insert (aa93–127), which is not shown in Figure 3(c) but is indicated by a circle. In spite of this insert/loop, the chimney left lip of scDNAP η resembles the left lip of hDNAP κ (Figure 3(c)/right) and of other Y-Family DNAPs. This insert serves as a "cap" over the chimney opening, such that the chimney is completely closed. DNAP η always has a cap, though its size varies (e.g., aa81–87 in hDNAP η). Speculation about a role for the DNAP η chimney cap is in the next section.

8. DNAP η Structures with TT-CPDs

Recently, X-ray structures of yeast DNAP η [39] and human DNAP η [40] with a TT-CPD were published, and remarkable insights have emerged. Two template bases are in the active site, with the base on the 3′-side base pairing with the dNTP. When the 3′-T of the TT-CPD interacts

with dATP, the 5′-T of the TT-CPD is also in the active site, and when the 5′-T of the TT-CPD is interacting with dATP, then the normal base on its 5′-side is in the active site. Undamaged DNA appears similarly. The T-bases of a TT-CPD lie at an angle of ∼30° with respect to each other and lack the usual twist between the base pairs; however, the impact of these distortions are minimized by the protein, such that the TT-CPD looks remarkably similar to a normal pair of adjacent thymines. Watson-Crick pairing is observed between the dATP and each T-base of the TT-CPD.

dATP adopts the canonical chair-like shape found in all families of DNAPs (see Section 10), though the shape is nuanced; for example, the angle of the A-base is tilted slightly downward compared to other dNTPs in Y-Family DNAPs in order to pair with the 3′-T of the TT-CPD. In hDNAP η, the guanidinium of R61 interacts with phosphate-oxygens on both the α- and β-positions of dATP, which is a unique interaction among Y-Family DNAPs. Interestingly, the equivalent R73 in yDNAP η is flexible and can be in this position, or it can face the opposite direction and pair with the extra template base in the active site; that is, the base not paired with the dNTP. This arginine is one of the most conserved amino acids in DNAP η, though the R73A mutation in yDNAP η retains normal kinetics with respect to both undamaged and damaged DNA, which suggests that its most important role is not being revealed in studies with a TT-CPD. In UmuC(V), which is the bacterial ortholog of DNAP η, methionine (M51) is usually at the equivalent position, though arginine or lysine are frequently present (42/408) (Figure 2). The other highly conserved amino acid in DNAP η is a glutamine (Q38 in hDNAP η and Q55 in yDNAP η), which sits in the minor groove and interacts with both O^2-positions on the T-bases of the TT-CPD.

Why is DNAP κ inactive on CPDs, while DNAP η is active? A number of structural elements no doubt contribute, but one important feature is that M135 in DNAP κ, which lies two positions before the roof-aa, is too bulky to accommodate both template-Ts of the TT-CPD. In the equivalent position, hDNAP η has a glycine (G46) and yDNAP η has a serine (S58), whose smaller size permits TT-CPD in the active site. Thus, DNAP κ may have an amino acid (i.e., M135) to minimize its activity on substrates meant for the Vη-class of DNAPs. Similarly, one of the functions of the flue and the cap for V/η-class DNAPs may be to minimize its activity on adducts that protrude into the minor groove, which are substrates for IV/κ-class DNAPs.

9. Architecture of the Y-Family Little Finger Domain

Y-Family DNAPs show considerable amino acid homology in the thumb/palm/fingers domains (approximately aa1–230), which makes alignment in this region, including for UmuC(V), unambiguous [44]. However, alignment of the little finger domain is more problematic. X-ray structures exist for the little finger domain of seven Y-Family DNAPs, and their fundamental structure is similar. They show a conserved secondary structure of β1-α1-β2-β3-α2-β4, where

the four β-strands are aligned and anti-parallel, while the two α-helices are aligned, antiparallel and cross-diagonally over the β-strands. In spite of this structural conservation, standard sequence alignment algorithms (e.g., ClustalW or MUSCLE) do not correctly align the little finger domains of these seven proteins.

Figure 4 shows the correct alignment based on the X-ray structures (as described in the legend), including the little finger domain of DNAP IV [112]. What features of these sequences allow the structures to be conserved, even though their primary amino acid sequences are not conserved? An inspection of the X-ray structures reveals that little finger domains are held together by a core of about twenty-one hydrophobic residues, which are highlighted in turquoise in the alignment in Figure 4 (H1–H21) and are shown in a Dpo4 structure (Figure 5). Though hydrophobicity is conserved at these twenty-one positions, the exact amino acid is not. Dpo4, Dbh, DNAP κ, DNAP IV, DNAP ι, yDNAP η, and hDNAP η have 19, 19, 17, 20, 20, 20, and 20 hydrophobic residues, respectively, at these 21 positions (Figure 4).

A comparison of these seven proteins reveals that the little finger domain has thirteen positions where an amino acid side-chain can interact with a phosphate-oxygen. Nine positions have a consensus lysine, arginine, asparagine or glutamine, which can interact with a phosphate-oxygen in the DNA backbone; they are designated L1–L9 in the alignment in Figure 4 (red) to indicate that their R-groups are "long." They are also shown in the Dpo4 structure in Figure 5 (red). Four positions have a consensus serine or threonine that can interact with a phosphate-oxygen; they are designated S1–S4 to indicate that their R-groups are "short" in Figure 4 (pink). They are also shown for Dpo4 in Figure 5 (pink).

In terms of DNA interactions, there are some nuances. In several cases amino acids with longer R-groups can also serve at the S1–S4 positions (e.g., K301 in Dbh). Regarding S3, S297/Dpo4, S297/Dbh and S359/DNAP ι clearly interact with the P + 6 phosphate-oxygen; however, T469/DNAP κ looks like a rotation would be required for it to interact properly, though it was counted as a positive. Q296 in DNAP IV might be able to interact with P + 5, though DNA is not present for definitive assessment and it is noncanonical, so it is not counted. R285/DNAP IV and R283/Dbh (instead of N340) might interact with P + 8, though DNA is not present for definitive assessment and it is non-canonical, so neither is counted.

Of the thirteen sites that can interact with phosphate-oxygens (i.e., L1–L9 and S1–S4), Dpo4 and Dbh have an appropriate amino acid at 13/13 sites and 11/13 sites, respectively. In contrast, hDNAP κ, DNAP IV, hDNAP ι, yDNAP η and hDNAP η have an appropriate amino acid at 9, 8, 7, 8, and 7 sites, respectively. The higher level of conformity for Dpo4 and Dbh undoubtedly reflects the need for more interactions with DNA given that they are from thermophilic bacteria and must operate at elevated temperatures. The similar number of residues ($\sim$8) for the other DNAPs probably reflects that they operate at a similar but lower temperature (i.e., 37°C), and if they had more

interactions they might bind DNA too tightly. We note that an increase in hydrophobic residues in the hydrophobic core was not expected for Dpo4 and Dbh, because hydrophobic interactions strengthen as temperature increases.

10. How Y-Family Architecture Influences dCTP versus dATP Insertion Opposite B[a]P Adducts

DNAP IV can pair dCTP with the dG moiety of +BP, importantly because the bulky pyrene can be accommodated in DNAP IV's large chimney opening (Figure 3(a)). For phosphoester bond formation to occur the distance between primer-O3′ and Pα-dCTP must be reaction-ready and can be compared to the closest possible distance, a van der Waals' contact ($\sim$3.5 Å). In models of +BP in DNAP IV [136], the distance between primer-O3′ and Pα-dCTP was $\sim$3.7 Å, which approximates a van der Waals' contact, and, thus, can be thought of as being "reaction ready." The no-adduct control had a similar primer-O3′ and Pα-dCTP distance ($\sim$3.7 Å).

In contrast, UmuC(V) does not give a satisfactory structure when +BP is paired with dCTP (Figure 3(b)/center), because UmuC(V)'s small chimney opening forces the bulky pyrene moiety downward. Asparagine-32 is the main problem, and its side chain plugs the UmuC(V) chimney leading to a clash with +BP. In the unadducted structure (Figure 3(b)/left), N32 adopts its lowest energy rotational conformer with respect to the Cα–Cβ bond. The presence of +BP leads to a rotation about the Cα–Cβ bond (Figure 3(b)/center); however, no other rotation can get N32 any farther out of the way. Consequently, UmuC(V)'s small chimney forces the +BP-dG in the template and its paired dCTP to move downward such that the primer-O3′/dCTP-Pα distance is elongated to $\sim$5.0 Å, which is a nonreaction-ready distance.

These observations provide a reasonable rationale for why DNAP IV preferentially does cellular dCTP insertion in cells: DNAP IV's large chimney opening permits a reasonable adduct-dG : dCTP structure with reaction-ready distances between primer-O3′ and Pα-dCTP.

If UmuC(V)'s small chimney enforces a non-reaction-ready distance between the primer-O3′/dNTP-Pα, then how could UmuC(V) insert any dNTP opposite +BP? Recently, we offered a hypothesis [136].

X-ray structures from all DNAP families show a canonical dNTP shape that has been called "chair-like," and its structure from T7 DNA polymerase is shown in Figure 6 (blue insert), which is also observed in most of the X-ray structures of Y-Family DNAPs, including Dpo4 ("S1-dNTP shape," Figure 6, green). However, a second non-canonical "goat-tail-like" shape ("S2-dNTP", Figure 6, yellow) has also been observed [26]. The goat-tail-like S2-dNTP shape can lie lower down in the active site, and +BP paired with dCTP in the S2-dNTP shape allows the pyrene to lie comfortably under the small chimney opening of UmuC(V), which allows the primer-O3′/dNTP-Pα distance to be reaction-ready ($\sim$3.8 Å).
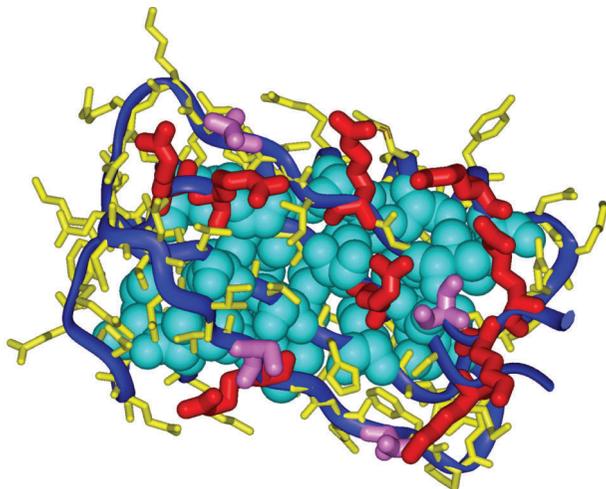
β1

	Conserved	L1	L2	S1	H1			L3		S2							
	Interaction site	T+3	T+5	T+4				T+2		T+1							
ssDpo4	242	R	K	S	I	G	R	I	V	T	M	K	R	-	-	N	S	R
ssDbh	243	K	I	P	H	G	G	R	Y	L	T	L	P	Y	-	-	N	T	R
hDNAP κ	413	R	K	S	M	G	V	E	R	T	F	S	-	-	-	E	I	N
DNAP IV	240	R	K	S	V	G	V	E	R	T	M	A	E	-	-	D	I	H
hDNAP ι	299	P	Q	S	F	S	M	E	E	D	S	F	K	K	-	-	-	C	S
yDNAPη	390	V	K	S	M	G	S	N	K	N	L	R	G	K	-	S	C	N
hDNAP η	316	P	K	T	I	G	C	S	K	N	F	P	G	K	T	A	L	A

α1

Interaction site (above): H2 ... H3 ... H4 H5 ... H6 ... L4 (T+2) H7

						H2					H3				H4	H5			H6				L4	H7													
ssDpo4	257	N	L	E	E	I	K	P	Y	L	F	R	A	I	E	E	S	Y	Y	K	L	D	K	R	-	-	-	-	-	-	-	-	-	I	P
ssDbh	258	D	V	K	V	I	L	P	Y	L	K	K	A	I	N	E	A	Y	Y	N	K	V	N	G	-	-	-	-	-	-	-	-	-	I	P
hDNAP κ	427	K	A	E	E	Q	Y	S	L	C	Q	E	L	C	S	E	L	A	Q	D	L	Q	K	E	R	-	-	-	-	-	-	L	K	G
DNAP IV	255	H	W	S	E	C	E	A	I	I	Q	E	R	L	Y	P	E	L	L	E	R	R	L	A	K	V	K	P	-	-	-	-	D	L	L	I
hDNAP ι	314	S	E	V	E	A	K	N	K	I	E	E	L	L	A	S	L	L	N	R	R	V	C	Q	D	G	R	K	P	-	-	P	R	K	P
yDNAPη	406	S	I	V	D	C	I	S	W	L	L	E	V	F	C	A	E	L	T	S	R	R	I	Q	D	L	E	Q	E	Y	N	K	I	V	I	P
hDNAP η	333	T	R	E	Q	V	Q	W	W	L	L	Q	L	A	Q	E	L	E	E	R	L	T	K	D	R	N	D	N	-	-	D	R	V	A

β2

				H8		H9		H10								
ssDpo4	282	K	A	I	H	V	V	A	V	T	E	-	-	-	-	D
ssDbh	282	M	R	I	T	V	I	A	I	M	E					D
hDNAP κ	454	R	T	V	T	I	K	L	K	N	V	-	-	-	-	N
DNAP IV	284	A	R	Q	G	V	K	L	K	F	D	-	-	-	-	D
hDNAP ι	340	H	T	V	R	L	I	I	R	R	Y	S	S	E	K	H
yDNAPη	439	R	T	V	S	I	S	L	K	T	K	-	-	-	-	S
hDNAP η	364	T	Q	L	V	V	S	I	R	V	Q	G	D	-	K	R

β3

Interaction site: S3 (P+6) L5 (P+6) ... S4 (P+8) ... H11

						S3	L5			S4														H11
ssDpo4	293	L	D	I	V	S	R	G	R	T	F	P	-	-	-	-	-	-	-	-	-	-	H	G	I
ssDbh	293	L	D	I	L	S	K	G	K	K	F	K	-	-	-	-	-	-	-	-	-	-	H	G	I
hDNAP κ	465	F	E	V	K	T	R	A	S	T	V	S	S	-	-	-	-	-	-	-	-	V	V	S
DNAP IV	295	F	Q	Q	T	T	Q	E	H	V	W	P	-	-	-	-	-	-	-	-	-	-	V	L
hDNAP ι	355	Y	G	R	E	S	R	Q	C	P	I	P	S	H	V	I	Q	K	L	G	T	G	N	Y	D
yDNAPη	450	Y	E	V	Y	R	K	S	G	P	V	A	Y	K	-	-	-	-	-	-	G	I	N	F
hDNAP η	378	L	S	S	L	R	R	C	C	A	L	T	-	-	-	-	-	-	-	-	-	R	Y

α2

										H12	H13			H14	H15			H16	H17									H18
ssDpo4	307	S	K	E	T	A	Y	S	E	S	V	K	L	L	Q	K	I	L	E	E	D	E	R	-	-	-	K	I
ssDbh	307	S	I	D	N	A	Y	K	V	A	E	D	L	L	L	R	E	L	L	V	R	D	K	R	R	-	N	V
hDNAP κ	480	T	A	E	E	I	F	A	I	A	K	E	L	L	K	T	E	I	D	A	D	F	P	H	P	-	L	R	L
DNAP IV	308	-	-	-	-	N	K	A	D	L	I	A	T	A	R	K	T	W	D	E	R	R	G	G	-	-	R	G	L
hDNAP ι	379	V	M	T	P	M	V	D	I	L	M	K	L	F	R	R	N	M	V	N	V	K	M	P	M	-	F	H	L
yDNAPη	468	Q	S	H	E	L	L	K	V	G	I	K	F	V	T	D	L	D	I	K	G	K	N	K	S	Y	Y	P	L
hDNAP η	391	D	A	H	K	M	S	H	D	A	F	T	V	I	K	N	C	N	T	S	G	I	Q	T	E	W	S	P	P	L

β4

Interaction site: L6 (T0) L7 (T+1) H19 ... H20 L8 (T+3) H21 ... L9 (P+8)

		L6	L7	H19		H20	L8	H21		L9		
ssDpo4	331	R	R	I	G	V	R	F	S	K	F	I
ssDbh	332	R	R	I	G	V	K	L	D	N	I	I
hDNAP κ	507	R	L	M	G	V	R	I	S	S	F	P
DNAP IV	330	R	L	V	G	L	H	V	T	L	L	D
hDNAP ι	405	T	L	L	S	V	C	F	C	N	L	K
yDNAPη	496	T	K	L	S	M	T	I	T	N	F	D
hDNAP η	420	T	M	L	F	L	C	A	T	K	F	S

FIGURE 4: Amino acid alignment for the little finger domains in cases where the alignment can be assigned unambiguously based on X-ray structures of Y-family DNAPs. Alignments were done with FATCAT [105] and MAMMOTH [106], which agreed with visual alignment. The six rows show the four β-strands (β1–β4) and the two α-helices (α1 and α2). Yellow highlighting indicates residues in each secondary structural block. Blue highlighting shows the twenty-one hydrophobic residues (H1–H21) that make up the hydrophobic core of the little finger domains. Red highlighting shows consensus lysine/arginine/asparagine/glutamine residues that bind with phosphate-oxygens (L1–L9). Pink highlighting shows consensus serine/threonine residues that bind with phosphate-oxygens (S1–S4). "Interaction Site" refers to the nucleotide position of the phosphate-oxygen with which an amino acid is interacting, where "T" refers to template and "P" refers to primer, and the numbering refers to the duplex portion of the DNA. The interaction site is not always the same in all structures in the case of amino acids with long side chains.

Interestingly, the S2-dATP shape allows the *syn*-conformation, such that *syn*-dATP can pair with adduct-dG via a Hoogsteen base pair. In contrast, the *syn*-adenine base in the S1-dNTP shape has steric clashes with atoms in the deoxyribose and the α-phosphate. Adduct-dG : *syn*-dATP pairing in UmuC(V) gave reasonable structures with primer-O3′/dNTP-Pα distances of ~3.6 Å [136]. The S2-dNTP shape appears to have accompanying protein components

that should allow phosphoester bond making and breaking [136, 138].

Other G : A mispairings are also possible. In principle, *anti*-dATP can pair with *syn*-guanine in adduct-dG, which requires the pyrene moiety to be in the major groove. *Anti*-dATP can also pair with *anti*guanine in adduct-dG in an elongated mispair. Thus, there are scenarios other than the one involving our *syn*-dATP : *anti*adduct-dG hypothesis.

FIGURE 5: The little finger domain of Dpo4, showing amino acids whose R-groups contribute to the hydrophobic core (H1–H21, turquoise). Furthermore, lysine/arginine/asparagine/glutamine residues that interact with phosphate-oxygens are shown in red (L1–L9), and serine/threonine residues that interact with phosphate-oxygens are shown in pink (S1–S4). The view is toward the four anti-parallel $\beta 1$–$\beta 4$ strands, and $\beta 1$ begins to the right, while the last amino acid in $\beta 4$ is shown just below it. The view is also into the face that binds duplex DNA, where the helix axis is approximately vertical. The Dpo4 coordinates are from 1SOM-B.
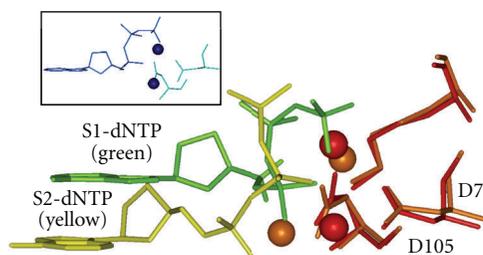
FIGURE 6: Side view of a dNTP in the "chair-like" shape S1-dNTP (green) versus the "goat-tail-like" shape S2-dNTP (yellow). Key amino acids (only D7, D105, and K159 are shown) from a Dpo4 structure adopting the S1-dNTP shape were superimposed on the same amino acids in a Dpo4 structure adopting the S2-dNTP. Spheres are divalent cations (S1-dNTP/red and S2-dNTP/brown). S1-dATP from T7 DNA polymerase is also shown (insert, blue). X-ray coordinates are from 1SOM-B for Dpo4/S1-dNTP, 1RYS-A for Dpo4/S2-dNTP, and 1T7P for T7 DNAP.

11. Unusual Architectural Features of Dpo4

Dpo4 is by far the best studied Y-Family DNAP, both structurally and biochemically. Based on biochemical and X-ray findings [28], Dpo4 insertion opposite +BP was proposed to follow a "dislocation" or "templated" pathway. Dislocation/templated insertion (see [135, 136] and references therein) involves DNAP stalling at an adduct, slippage to the next 5′-base along the template, which directs incorporation (e.g., dATP insertion opposite the 5′-T in a 5′-TG sequence context), whereupon the newly incorporated

dA slips back to form an adduct-G : A mispair, from which extension yields the mispair that ultimately gives a G->T mutation. Dpo4 preferentially inserted dCTP, dTTP, dATP and dGTP opposite +BP in 5′-GG, 5′-AG, 5′-TG and 5′-CG sequences, respectively, [28], which is consistent with a dislocation/templated mechanism.

Though the dislocation/templated mechanism is attractive for Dpo4, considerable evidence both in vitro and in vivo suggest that neither DNAP IV nor DNAP V follow a dislocation/templated mechanism with +BP, as discussed extensively in [136] and references therein.

Why might Dpo4 be different than DNAPs IV and V? Dpo4 is in the IV/κ-class, and it has a nonbulky roof-aa and roof-neighbor-aa [A44/A57], as expected for the IV/κ-class. However, Dpo4 has a very small chimney opening (discussed below), which is associated with for the V/η-class. Thus, Dpo4 is a hybrid with a roof similar to the IV/κ-class and a chimney similar to the V/η-class.

Dpo4 has a small chimney opening, because its bulky flue-handle (C31) causes downward curvature of the chimney upper lip and leads to a closed flue (V32) [136]. In fact, Dpo4's chimney is exceedingly blocked: (1) the V32 flue is inserted deeper into the chimney than, for example, the N32 flue of UmuC(V), and (2) M76, which is the second amino acid in Dpo4's left lip, also plugs the chimney. DNAP IV and UmuC(V) have non-bulky G74 and S72, respectively, in the position equivalent to M76 in Dpo4. Thus, the excessively plugged chimney of Dpo4 forces the pyrene moiety of +BP so far from the active site that base pairing via either S1-dCTP or S2-dCTP is impossible; consequently, both the pyrene and the dG moieties of +BP are forced to be extrahelical with consequence being that pairing cannot occur with its complementary dC [28].

As mentioned above, DNAPs η, IV, η, and ι have an appropriate amino acid at 9, 8, 7, and 8, respectively, of the thirteen sites that can interact with phosphate-oxygens (L1–L9 and S1–S4). In contrast, Dpo4 and Dbh conform 13/13 and 11/13, respectively, which undoubtedly reflects the need for more interactions with DNA given that they are from thermophilic bacteria and must operate at an elevated temperature. Thus, Dpo4 studied at 37°C, which is typical, may give results that do not reflect correctly on aspects of the mechanism of other Y-Family DNAPs, which have evolved to operate at 37°C.

This analysis suggests reasons for caution when applying conclusions from Dpo4 to other Y-Family DNAPs, especially those purely in the IV/κ-class or the V/η-class. Perhaps Dpo4 evolved its hybrid roof/chimney structure to bypass a unique set of lesions encountered by a thermophilic bacteria. Alternatively, perhaps the structure of Dpo4 at physiologically relevant elevated temperatures is different than at the temperature at which it was crystallized (room temperature.) and assayed (37°C), and this affects its structure and behavior.

12. Structure of B-Family Lesion-Bypass DNAPs

This paper has focused on Y-Family DNAPs, though some lesion bypass DNAPs are in the B-Family, including DNAP

II in *E. coli* and REV3 of DNAP ξ in many eukaryotic cells. DNAP II inserts and extends the -2 frameshift intermediate of AAF-C8-dG [66, 117], which must have two looped out nucleotides as well as the AAF moiety protruding into the major groove. Data also suggests that DNAP II and DNAP ξ are involved in the bypass of interstrand crosslinks [45–49], which must have a large oligonucleotide protruding into the major groove during TLS. Though B-Family DNAPs completely surround DNA, the structure on the minor and major groove sides are very different, as revealed in structures of both *E. coli* DNAP II [139] and Rb69 DNAP [140]. B-family DNAPs have a helical protein component that follows and contacts the minor groove side of duplex DNA. On the major grove side, however, a protein dome is present that leaves a large open cavity. Though Y-Family DNAPs are open to solvent on their major groove side, the solvent-exposed DNA surface inside the cavity for DNAP II ($\sim 400\,\text{Å}^2$, when considering, for example, the template : dNTP base pair plus the L + 1 base pair) is actually larger than with either DNAP IV ($\sim 230\,\text{Å}^2$) or UmuC(V) ($\sim 130\,\text{Å}^2$). It seems likely that the large cavity and solvent exposed region on the major groove side of B-Family DNAPs may be essential for their ability to accomplish TLS on lesions having bulky protrusions into the major groove.

Abbreviations

B[a]P:	Benzo[a]pyrene
+BP:	[+ta]-B[a]P-N^2-dG (Figure 1)
−BP:	[−ta]-B[a]P-N^2-dG
TT-CPD:	Thymine-thymine cyclopyrimidine dimer
TLS:	Translesion synthesis, which includes the insertion of a base opposite a DNA adduct, as well as subsequent elongation
DNAP:	DNA polymerase
S1-dNTP:	The "chair-like" dNTP shape
S2-dNTP:	The "goat-tail-like" dNTP shape
aa:	Amino acid.

References

[1] R. A. Weinberg, *The Biology of Cancer*, Garland Science, Taylor & Francis, New York, NY, USA, 2007.

[2] A. Luch, "Nature and nurture—lessons from chemical carcinogenesis," *Nature Reviews Cancer*, vol. 5, no. 2, pp. 113–125, 2006.

[3] B. Singer and D. Grunberger, *Molecular Biology of Mutagens and Carcinogens*, Plenum Press, New York, NY, USA, 1983.

[4] A. Balmain, R. Brown, and C. C. Harris, Eds., "The twentieth anniversary special issue contains reviews of many carcinogenesis topics," *Carcinogenesis*, vol. 21, pp. 339–531, 2000.

[5] A. H. Conney, "Induction of microsomal enzymes for foreign chemicals and carcinogenesis by polycyclic aromatic hydrocarbons: G.H.A. Clowes Memorial Lecture," *Cancer Research*, vol. 42, no. 12, pp. 4875–4917, 1982.

[6] S. D. McCulloch and T. A. Kunkel, "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases," *Cell Research*, vol. 18, no. 1, pp. 148–161, 2008.

[7] S. D. McCulloch and T. A. Kunkel, "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases," *Cell Research*, vol. 18, no. 1, pp. 148–161, 2008.

[8] C. Guo, J. N. Kosarek-Stancel, T.-S. Tang, and E. C. Friedberg, "Y-family DNA polymerases in mammalian cells," *Cellular and Molecular Life Sciences*, vol. 66, no. 14, pp. 2363–2381, 2009.

[9] L. S. Waters, B. K. Minesinger, M. E. Wiltrout, S. D'Souza, R. V. Woodruff, and G. C. Walker, "Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance," *Microbiology and Molecular Biology Reviews*, vol. 73, no. 1, pp. 134–154, 2009.

[10] J. D. Pata, "Structural diversity of the Y-family DNA polymerases," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1124–1135, 2010.

[11] S. Schneider, S. Schorr, and T. Carell, "Crystal structure analysis of DNA lesion repair and tolerance mechanisms," *Current Opinion in Structural Biology*, vol. 19, no. 1, pp. 87–95, 2009.

[12] M. T. Washington, K. D. Carlson, B. D. Freudenthal, and J. M. Pryor, "Variations on a theme: eukaryotic Y-family DNA polymerases," *Biochimica et Biophysica Acta*, vol. 1804, no. 5, pp. 1113–1123, 2010.

[13] K. Bebenek and T. A. Kunkel, "Functions of DNA polymerases," *Advances in Protein Chemistry*, vol. 69, pp. 137–165, 2004.

[14] P. J. Rothwell and G. Waksman, "Structure and mechanism of DNA polymerases," *Advances in Protein Chemistry*, vol. 71, pp. 401–440, 2005.

[15] E. C. Friedberg, G. W. Walker, W. Siede, R. D. Wood, R. A. Schultz, and T. Ellenberger, *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, DC, USA, 2nd edition, 2006.

[16] W. Yang and R. Woodgate, "What a difference a decade makes: insights into translesion DNA synthesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 40, pp. 15591–15598, 2007.

[17] H. Ohmori, E. C. Friedberg, R. P. P. Fuchs et al., "The Y-family of DNA polymerases," *Molecular Cell*, vol. 8, no. 1, pp. 7–8, 2001.

[18] T. Nohmi, "Environmental stress and lesion-bypass DNA polymerases," *Annual Review of Microbiology*, vol. 60, pp. 231–253, 2006.

[19] W. Yang, "Damage repair DNA polymerases Y," *Current Opinion in Structural Biology*, vol. 13, no. 1, pp. 23–30, 2003.

[20] S. Prakash, R. E. Johnson, and L. Prakash, "Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function," *Annual Review of Biochemistry*, vol. 74, pp. 317–353, 2005.

[21] D. F. Jarosz, P. J. Beuning, S. E. Cohen, and G. C. Walker, "Y-family DNA polymerases in Escherichia coli," *Trends in Microbiology*, vol. 15, no. 2, pp. 70–77, 2007.

[22] R. P. Fuchs, S. Fujii, and J. Wagner, "Properties and functions of Escherichia coli: Pol IV and Pol V," *Advances in Protein Chemistry*, vol. 69, pp. 229–264, 2004.

[23] B.-L. Zhou, J. D. Pata, and T. A. Steitz, "Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain," *Molecular Cell*, vol. 8, no. 2, pp. 427–437, 2001.

[24] H. Ling, F. Boudsocq, R. Woodgate, and W. Yang, "Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication," *Cell*, vol. 107, no. 1, pp. 91–102, 2001.

[25] H. Ling, F. Boudsocq, B. S. Plosky, R. Woodgate, and W. Yang, "Replication of a cis-syn thymine dimer at atomic resolution," *Nature*, vol. 424, no. 6952, pp. 1083–1087, 2003.

[26] A. Vaisman, H. Ling, R. Woodgate, and W. Yang, "Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis," *EMBO Journal*, vol. 24, no. 17, pp. 2957–2967, 2005.

[27] O. Rechkoblit, L. Malinina, Y. Cheng et al., "Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion," *PLoS Biology*, vol. 4, no. 1, pp. 25–42, 2006.

[28] J. Bauer, G. Xing, H. Yagi, J. M. Sayer, D. M. Jerina, and H. Ling, "A structural gap in Dpo4 supports mutagenic bypass of a major benzo [a]pyrene dG adduct in DNA through template misalignment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 38, pp. 14905–14910, 2007.

[29] J. H. Wong, K. A. Fiala, Z. Suo, and H. Ling, "Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4," *Journal of Molecular Biology*, vol. 379, no. 2, pp. 317–330, 2008.

[30] R. L. Eoff, R. Sanchez-Ponce, and F. P. Guengerich, "Conformational changes during nucleotide selection by Sulfolobus solfataricus DNA polymerase Dpo4," *Journal of Biological Chemistry*, vol. 284, no. 31, pp. 21090–21099, 2009.

[31] H. Zhang and F. P. Guengerich, "Effect of N^2-guanyl modifications on early steps in catalysis of polymerization by Sulfolobus solfataricus P2 DNA polymerase Dpo4 T239W," *Journal of Molecular Biology*, vol. 395, no. 5, pp. 1007–1018, 2010.

[32] O. Rechkoblit, A. Kolbanovskiy, L. Malinina, N. E. Geacintov, S. Broyde, and D. J. Patel, "Mechanism of error-free and semitargeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4," *Nature Structural and Molecular Biology*, vol. 17, no. 3, pp. 379–388, 2010.

[33] J. Trincao, R. E. Johnson, C. R. Escalante, S. Prakash, L. Prakash, and A. K. Aggarwal, "Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis," *Molecular Cell*, vol. 8, no. 2, pp. 417–426, 2001.

[34] A. Alt, K. Lammens, C. Chiocchini et al., "Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase η," *Science*, vol. 318, no. 5852, pp. 967–970, 2007.

[35] S. Lone, S. A. Townson, S. N. Uljon et al., "Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass," *Molecular Cell*, vol. 25, no. 4, pp. 601–614, 2007.

[36] D. T. Hair, R. E. Johnson, S. Prakash, L. Prakash, and A. K. Aggarwal, "Replication by human DNA polymerase-ι occurs by Hoogsteen base-pairing," *Nature*, vol. 430, no. 6997, pp. 377–380, 2004.

[37] D. T. Nair, R. E. Johnson, L. Prakash, S. Prakash, and A. K. Aggarwal, "Human DNA polymerase ι incorporates dCTP opposite template G via a G.C+ hoogsteen base pair," *Structure*, vol. 13, no. 10, pp. 1569–1577, 2005.

[38] D. T. Nair, R. E. Johnson, L. Prakash, S. Prakash, and A. K. Aggarwal, "Hoogsteen base pair formation promotes synthesis opposite the 1,N 6-ethenodeoxyadenosine lesion by human DNA polymerase ι," *Nature Structural and Molecular Biology*, vol. 13, no. 7, pp. 619–625, 2006.

[39] T. D. Silverstein, R. E. Johnson, R. Jain, L. Prakash, S. Prakash, and A. K. Aggarwal, "Structural basis for the suppression of skin cancers by DNA polymerase," *Nature*, vol. 465, no. 7301, pp. 1039–1043, 2010.

[40] C. Biertümpfel, Y. Zhao, Y. Kondo et al., "Structure and mechanism of human DNA polymerase," *Nature*, vol. 465, no. 7301, pp. 1044–1048, 2010.

[41] R. E. Johnson, S. Prakash, and L. Prakash, "The human DINB1 gene encodes the DNA polymerase Polθ," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 8, pp. 3838–3843, 2000.

[42] E. Ohashi, T. Ogi, R. Kusumoto et al., "Error-prone bypass of certain DNA lesions by the human DNA polymerase κ," *Genes and Development*, vol. 14, no. 13, pp. 1589–1594, 2000.

[43] Y. Zhang, F. Yuan, X. Wu et al., "Error-free and error-prone lesion bypass by human DNA polymerase κ in vitro," *Nucleic Acids Research*, vol. 28, no. 21, pp. 4138–4146, 2000.

[44] C. H. Lee, S. Chandani, and E. L. Loechler, "Homology modeling of four Y-family, lesion-bypass DNA polymerases: the case that E. coli Pol IV and human Pol κ are orthologs, and E. coli Pol V and human Pol η are orthologs," *Journal of Molecular Graphics and Modelling*, vol. 25, no. 1, pp. 87–102, 2006.

[45] M. Berardini, W. Mackay, and E. L. Loechler, "A site-specific study of a plasmid containing single nitrogen mustard interstrand cross-link: evidence for a second, recombination-independent pathway for the DNA repair of interstrand cross-links," *Biochemistry*, vol. 36, pp. 3506–3513, 1997.

[46] S. Sarkar, A. A. Davies, H. D. Ulrich, and P. J. McHugh, "DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase ζ," *EMBO Journal*, vol. 25, no. 6, pp. 1285–1294, 2006.

[47] P. J. McHugh and S. Sarkar, "DNA interstrand cross-link repair in the cell cycle: a critical role for polymerase ζ in G1 phase," *Cell Cycle*, vol. 5, no. 10, pp. 1044–1047, 2006.

[48] P. Lehoczký, P. J. McHugh, and M. Chovanec, "DNA interstrand cross-link repair in Saccharomyces cerevisiae," *FEMS Microbiology Reviews*, vol. 31, no. 2, pp. 109–133, 2007.

[49] M. Räschle, P. Knipscheer, M. Enoiu et al., "Mechanism of replication-coupled DNA interstrand crosslink repair," *Cell*, vol. 134, no. 6, pp. 969–980, 2008.

[50] K. Schlacher and M. F. Goodman, "Lessons from 50 years of SOS DNA-damage-induced mutagenesis," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 7, pp. 587–594, 2000.

[51] Z. Qiu and M. F. Goodman, "The Escherichia coli polB locus is identical to dinA, the structural gene for DNA polymerase II: characterization of pol II purified from A polB mutant," *Journal of Biological Chemistry*, vol. 272, no. 13, pp. 8611–8617, 1997.

[52] S.-R. Kim, K. Matsui, M. Yamada, P. Gruz, and T. Nohmi, "Roles of chromosomal and episomal dinB genes encoding DNA pol IV in targeted and untargeted mutagenesis in *Escherichia coli*," *Molecular Genetics and Genomics*, vol. 266, no. 2, pp. 207–215, 2001.

[53] R. Woodgate and D. G. Ennis, "Levels of chromosomally encoded Umu proteins and requirements for in vivo UmuD cleavage," *Molecular and General Genetics*, vol. 229, no. 1, pp. 10–16, 1991.

[54] M. Tang, P. Pham, X. Shen et al., "Roles of E. coli DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis," *Nature*, vol. 404, no. 6781, pp. 1014–1018, 2000.

[55] D. E. Brash and W. A. Haseltine, "UV-induced mutation hotspots occur at DNA damage hotspots," *Nature*, vol. 298, no. 5870, pp. 189–192, 1982.

[56] M. F. Goodman, "Error-prone repair DNA polymerases in prokaryotes and eukaryotes," *Annual Review of Biochemistry*, vol. 71, pp. 17–50, 2002.

[57] J. Yin, K. Y. Seo, and E. L. Loechler, "A role for DNA polymerase V in G→T mutations from the major benzo[a]pyrene N^2-dG adduct when studied in a 5′-TGT sequence in E. coli," *DNA Repair*, vol. 3, no. 3, pp. 323–334, 2004.

[58] B. S. Strauss, "The "A" rule revisited: polymerases as determinants of mutational specificity," *DNA Repair*, vol. 1, no. 2, pp. 125–135, 2002.

[59] E. L. Loechler, "Mechanism by which aflatoxins and other bulky carcinogens induce mutations," in *The Toxicology of Aflatoxins: Human Health, Veterinary, and Agricultural Significance*, D. L. Eaton and J. D. Groopman, Eds., chapter 8, pp. 149–178, Academic Press, Orlando, Fla, USA, 1994.

[60] C. Masutani, R. Kusumoto, A. Yamada et al., "The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η," *Nature*, vol. 399, no. 6737, pp. 700–704, 1999.

[61] R. E. Johnson, S. Prakash, and L. Prakash, "Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta," *Science*, vol. 283, no. 5404, pp. 1001–1004, 1999.

[62] R. E. Johnson, C. M. Kondratick, S. Prakash, and L. Prakash, "hRAD30 mutations in the variant form of xeroderma pigmentosum," *Science*, vol. 285, no. 5425, pp. 263–265, 1999.

[63] M. T. Washington, R. E. Johnson, S. Prakash, and L. Prakash, "Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase η," *Journal of Biological Chemistry*, vol. 274, no. 52, pp. 36835–36838, 1999.

[64] M. T. Washington, R. E. Johnson, S. Prakash, and L. Prakash, "Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase η," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 7, pp. 3094–3099, 2000.

[65] R. E. Johnson, M. T. Washington, S. Prakash, and L. Prakash, "Fidelity of human DNA polymerase η," *Journal of Biological Chemistry*, vol. 275, no. 11, pp. 7447–7450, 2000.

[66] S. Fujii and R. P. Fuchs, "Interplay among replicative and specialized DNA polymerases determines failure or success of translesion synthesis pathways," *Journal of Molecular Biology*, vol. 372, no. 4, pp. 883–893, 2007.

[67] Q. Jiang, K. Karata, R. Woodgate, M. M. Cox, and M. F. Goodman, "The active form of DNA polymerase v is UmuD′2 C-RecA-ATP," *Nature*, vol. 460, no. 7253, pp. 359–363, 2009.

[68] M. Patel, Q. Jiang, R. Woodgate, M. M. Cox, and M. F. Goodman, "A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 45, no. 3, pp. 171–184, 2010.

[69] S. Fujii, A. Isogawa, and R. P. Fuchs, "RecFOR proteins are essential for Pol V-mediated translesion synthesis and mutagenesis," *EMBO Journal*, vol. 25, no. 24, pp. 5754–5763, 2006.

[70] M. J. McIlwraith, A. Vaisman, Y. Liu, E. Fanning, R. Woodgate, and S. C. West, "Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination," *Molecular Cell*, vol. 20, pp. 783–792, 2005.

[71] S. Kobayashi, M. R. Valentine, P. Pham, M. O'Donnell, and M. F. Goodman, "Fidelity of Escherichia coli DNA polymerase IV. Preferential generation of small deletion mutations by dNTP-stabilized misalignment," *Journal of Biological Chemistry*, vol. 277, no. 37, pp. 34198–34207, 2002.

[72] D. F. Jarosz, V. G. Godoy, J. C. Delaney, J. M. Essigmann, and G. C. Walker, "A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates," *Nature*, vol. 439, no. 7073, pp. 225–228, 2006.

[73] N. Lenne-Samuel, R. Janel-Bintz, A. Kolbanovskiy, N. E. Geacintov, and R. P. P. Fuchs, "The processing of a Benzo(a)pyrene adduct into a frameshift or a base substitution mutation requires a different set of genes in Escherichia coli," *Molecular Microbiology*, vol. 38, no. 2, pp. 299–307, 2000.

[74] R. Napolitano, R. Janel-Bintz, J. Wagner, and R. P. P. Fuchs, "All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis," *EMBO Journal*, vol. 19, no. 22, pp. 6259–6265, 2000.

[75] K. Y. Seo, A. Nagalingam, S. Miri et al., "Mirror image stereoisomers of the major benzo[a]pyrene N^2-dG adduct are bypassed by different lesion-bypass DNA polymerases in E. coli," *DNA Repair*, vol. 5, no. 4, pp. 515–522, 2006.

[76] B. Yuan, H. Cao, Y. Jiang, H. Hong, and Y. Wang, "Efficient and accurate bypass of N^2-(1-carboxyethyl)-2′-deoxyguanosine by DinB DNA polymerase in vitro and in vivo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 25, pp. 8679–8684, 2008.

[77] X. Shen, J. M. Sayer, H. Kroth et al., "Efficiency and accuracy of SOS-induced DNA polymerases replicating benzo[a]pyrene-7,8-diol 9,10-epoxide A and G adducts," *Journal of Biological Chemistry*, vol. 277, no. 7, pp. 5265–5274, 2002.

[78] L. J. Marnett, "Oxyradicals and DNA damage," *Carcinogenesis*, vol. 21, no. 3, pp. 361–370, 2000.

[79] T. Opperman, S. Murli, B. T. Smith, and G. C. Walker, "A model for a umuDC-dependent prokaryotic DNA damage checkpoint," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 16, pp. 9218–9223, 1999.

[80] C. Indiani, L. D. Langston, O. Yurieva, M. F. Goodman, and M. O'Donnell, "Translesion DNA polymerases remodel the replisome and alter the speed of the replicative helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 15, pp. 6031–6038, 2009.

[81] B. Yeiser, E. D. Pepper, M. F. Goodman, and S. E. Finkel, "SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 13, pp. 8737–8741, 2002.

[82] J. D. Tompkins, J. L. Nelson, J. C. Hazel, S. L. Leugers, J. D. Stumpf, and P. L. Foster, "Error-prone polymerase, DNA polymerase IV, is responsible for transient hypermutation during adaptive mutation in Escherichia coli," *Journal of Bacteriology*, vol. 185, no. 11, pp. 3469–3472, 2003.

[83] D. T. Nair, R. E. Johnson, L. Prakash, S. Prakash, and A. K. Aggarwal, "Rev1 employs a novel mechanism of DNA synthesis using a protein template," *Science*, vol. 309, no. 5744, pp. 2219–2222, 2006.

[84] S. Avkin, M. Goldsmith, S. Velasco-Miguel, N. Geacintov, E. C. Friedberg, and Z. Livneh, "Quantitative analysis of translesion DNA synthesis across a benzo[a]pyrene-guanine adduct in mammalian cells: the role of DNA polymerase," *Journal of Biological Chemistry*, vol. 279, no. 51, pp. 53298–53305, 2004.

[85] T. Ogi, Y. Shinkai, K. Tanaka, and H. Ohmori, "Polκ protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 24, pp. 15548–15553, 2002.

[86] O. Rechkoblit, Y. Zhang, D. Guo et al., "trans-lesion synthesis past bulky benzo[a]pyrene diol epoxide N^2-dG and N6-dA lesions catalyzed by DNA bypass polymerases," *Journal of Biological Chemistry*, vol. 277, no. 34, pp. 30488–30494, 2002.

[87] L. Jia, N. E. Geacintov, and S. Broyde, "The N-clasp of human DNA polymerase κ promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions," *Nucleic Acids Research*, vol. 36, no. 20, pp. 6571–6584, 2008.

[88] S. D. McCulloch, A. Wood, P. Garg, P. M. J. Burgers, and T. A. Kunkel, "Effects of accessory proteins on the bypass of a cis-syn thymine-thymine dimer by Saccharomyces cerevisiae DNA polymerase η," *Biochemistry*, vol. 46, no. 30, pp. 8888–8896, 2007.

[89] S. D. McCulloch, R. J. Kokoska, C. Masutani, S. Iwai, F. Hanaoka, and T. A. Kunkel, "Preferential cis-syn thymine dimer bypass by DNA polymerase η occurs with biased fidelity," *Nature*, vol. 428, no. 6978, pp. 97–100, 2004.

[90] L. Haracska, S.-L. Yu, R. E. Johnson, L. Prakash, and S. Prakash, "Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η," *Nature Genetics*, vol. 25, no. 4, pp. 458–461, 2000.

[91] A. Vaisman, C. Masutani, F. Hanaoka, and S. G. Chaney, "Efficient translesion replication past oxaliplatin and cisplatin GpG adducts by human DNA polymerase η," *Biochemistry*, vol. 39, no. 16, pp. 4575–4580, 2000.

[92] T. B. Petta, S. Nakajima, A. Zlatanou et al., "Human DNA polymerase iota protects cells against oxidative stress," *EMBO Journal*, vol. 27, no. 21, pp. 2883–2895, 2008.

[93] Y. Zhang, F. Yuan, X. Wu, J.-S. Taylor, and Z. Wang, "Response of human DNA polymerase ι to DNA lesions," *Nucleic Acids Research*, vol. 29, no. 4, pp. 928–935, 2001.

[94] K. Donny-Clark, R. Shapiro, and S. Broyde, "Accommodation of an N-(deoxyguanosin-8-yl)-2-acetylaminofluorene adduct in the active site of human DNA polymerase ι: Hoogsteen or Watson−Crick base pairing?" *Biochemistry*, vol. 48, no. 1, pp. 7–18, 2009.

[95] L. Wang and S. Broyde, "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 Research*, vol. 34, no. 3, pp. 785–795, 2006.

[96] C. Xu, B. A. Maxwell, J. A. Brown, L. Zhang, and Z. Suo, "Global conformational dynamics of a Y-family DNA polymerase during catalysis," *PLoS Biology*, vol. 7, no. 10, Article ID e1000225, 2009.

[97] M. A. Booher, S. Wang, and E. T. Kool, "Base pairing and steric interactions between pyrimidine strand bridging loops and the purine strand in DNA pyrimidine-purine-pyrimidine triplexes," *Biochemistry*, vol. 33, no. 15, pp. 4645–4651, 1994.

[98] M. T. Washington, S. A. Helquist, E. T. Kool, L. Prakash, and S. Prakash, "Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase η," *Molecular and Cellular Biology*, vol. 23, no. 14, pp. 5107–5112, 2003.

[99] W. T. Wolfle, M. T. Washington, E. T. Kool et al., "Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase κ," *Molecular and Cellular Biology*, vol. 25, no. 16, pp. 7137–7143, 2005.

[100] L. Wang, X. Yu, P. Hu, S. Broyde, and Y. Zhang, "A water-mediated and substrate-assisted catalytic mechanism for Sulfolobus solfataricus DNA polymerase IV," *Journal of the American Chemical Society*, vol. 129, no. 15, pp. 4731–4737, 2007.

[101] I. Letunic, L. Goodstadt, N. J. Dickens et al., "Recent improvements to the SMART domain-based sequence-annotation resource," *Nucleic Acids Research*, vol. 30, no. 1, pp. 242–244, 2002.

[102] A. Bairoch, R. Apweiler, C. H. Wu et al., "The universal protein resource (UniProt)," *Nucleic Acids Research*, vol. 33, pp. D154–D159, 2005.

[103] T. Lima, A. H. Auchincloss, E. Coudert et al., "HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot," *Nucleic Acids Research*, vol. 37, no. 1, pp. D471–D478, 2009.

[104] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[105] Y. Ye and A. Godzik, "Flexible structure alignment by chaining aligned fragment pairs allowing twists," *Bioinformatics*, vol. 19, supplement 2, pp. ii246–ii255, 2003.

[106] A. R. Ortiz, C. E. M. Strauss, and O. Olmea, "MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison," *Protein Science*, vol. 11, no. 11, pp. 2606–2621, 2002.

[107] S. Fujii and R. P. Fuchs, "Biochemical basis for the essential genetic requirements of RecA and the β-clamp in Pol V activation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 35, pp. 14825–14830, 2009.

[108] R. C. Heller and K. J. Marians, "Replisome assembly and the direct restart of stalled replication forks," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 12, pp. 932–943, 2006.

[109] L. D. Langston and M. O'Donnell, "DNA replication: keep moving and don't mind the gap," *Molecular Cell*, vol. 23, no. 2, pp. 155–160, 2006.

[110] A. Furukohri, M. F. Goodman, and H. Maki, "A dynamic polymerase exchange with Escherichia coli DNA polymerase IV replacing DNA polymerase III on the sliding clamp," *Journal of Biological Chemistry*, vol. 283, no. 17, pp. 11260–11269, 2008.

[111] O. J. Becherel, R. P. P. Fuchs, and J. Wagner, "Pivotal role of the β-clamp in translesion DNA synthesis and mutagenesis in E. coli cells," *DNA Repair*, vol. 4, no. 9, pp. 703–708, 2002.

[112] K. A. Bunting, S. M. Roe, and L. H. Pearl, "Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the β-clamp," *EMBO Journal*, vol. 22, no. 21, pp. 5883–5892, 2003.

[113] J. M. H. Heltzel, R. W. Maul, S. K. Scouten Ponticelli, and M. D. Sutton, "A model for DNA polymerase switching involving a single cleft and the rim of the sliding clamp," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 31, pp. 12664–12669, 2009.

[114] J. M. H. Heltzel, S. K. Scouten Ponticelli, L. H. Sanders et al., "Sliding clamp-DNA interactions are required for viability and contribute to DNA polymerase management in Escherichia coli," *Journal of Molecular Biology*, vol. 387, no. 1, pp. 74–91, 2009.

[115] J. Wagner, H. Etienne, R. P. Fuchs, A. Cordonnier, and D. Burnouf, "Distinct β-clamp interactions govern the activities of the y family PolIV DNA polymerase," *Molecular Microbiology*, vol. 74, no. 5, pp. 1143–1151, 2009.

[116] S. Delmas and I. Matic, "Interplay between replication and recombination in Escherichia coli: impact of the alternative DNA polymerases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 12, pp. 4564–4569, 2006.

[117] R. P. Fuchs and S. Fujii, "Translesion synthesis in Escherichia coli: lessons from the NarI mutation hot spot," *DNA Repair*, vol. 6, no. 7, pp. 1032–1041, 2007.

[118] R. H. Heflich and R. E. Neft, "Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites," *Mutation Research*, vol. 318, no. 2, pp. 73–174, 1994.

[119] L. Wang and S. Broyde, "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 Research*, vol. 34, no. 3, pp. 785–795, 2006.

[120] R. G. Harvey, *Polycyclic Aromatic Hydrocarbons: Chemistry and Cancer*, Wiley-VCH, New York, NY, USA, 1997.

[121] A. Dipple, "Polycyclic aromatic hydrocarbon carcinogens," in *Polycyclic Aromatic Hydrocarbons and Carcinogenesis*, R. G. Harvey, Ed., pp. 1–17, American Chemical Society Press, Washington, DC, USA, 1985.

[122] M. F. Denissenko, A. Pao, M.-S. Tang, and G. P. Pfeifer, "Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53," *Science*, vol. 274, no. 5286, pp. 430–432, 1996.

[123] M. F. Denissenko, A. Pao, G. P. Pfeifer, and M.-S. Tang, "Slow repair of bulky DNA adducts along the nontranscribed strand of the human p53 gene may explain the strand bias of transversion mutations in cancers," *Oncogene*, vol. 16, no. 10, pp. 1241–1247, 1998.

[124] M.-S. Tang, J. B. Zheng, M. F. Denissenko, G. P. Pfeifer, and Y. Zheng, "Use of UvrABC nuclease to quantify benzo[a]pyrene diol epoxide-DNA adduct formation at methylated versus unmethylated CpG sites in the p53 gene," *Carcinogenesis*, vol. 20, no. 6, pp. 1085–1089, 1999.

[125] G. P. Pfeifer and M. F. Denissenko, "Formation and repair of DNA lesions in the p53 gene: relation to cancer mutations?" *Environmental and Molecular Mutagenesis*, vol. 31, no. 3, pp. 197–205, 1998.

[126] G. P. Pfeifer, M. F. Denissenko, M. Olivier, N. Tretyakova, S. S. Hecht, and P. Hainaut, "Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers," *Oncogene*, vol. 21, no. 48, pp. 7435–7451, 2002.

[127] L. E. Smith, M. F. Denissenko, W. P. Bennett et al., "Targeting of lung cancer mutational hotspots by polycyclic aromatic hydrocarbons," *Journal of the National Cancer Institute*, vol. 92, no. 10, pp. 803–811, 2000.

[128] G. P. Pfeifer and P. Hainaut, "On the origin of $G \rightarrow T$ transversions in lung cancer," *Mutation Research*, vol. 526, no. 1-2, pp. 39–43, 2003.

[129] H. Rodriguez and E. L. Loechler, "Mutagenesis by the (+)-anti-diol epoxide of benzo[a]pyrene: what controls mutagenic specificity?" *Biochemistry*, vol. 32, no. 7, pp. 1759–1769, 1993.

[130] Z. Xie, E. Braithwaite, D. Guo, B. Zhao, N. E. Geacintov, and Z. Wang, "Mutagenesis of benzo[a]pyrene diol epoxide in yeast: requirement for DNA polymerase ζ and involvement of DNA polymerase η," *Biochemistry*, vol. 42, no. 38, pp. 11253–11262, 2003.

[131] J.-H. Yoon, C.-S. Lee, and G. P. Pfeifer, "Simulated sunlight and benzo[a]pyrene diol epoxide induced mutagenesis in the human p53 gene evaluated by the yeast function assay: lack of correspondence to tumor mutation spectra," *Carcinogenesis*, vol. 24, no. 1, pp. 113–119, 2003.

[132] M. Schiltz, X. X. Cui, Y.-P. Lu et al., "Characterization of the mutational profile of (+)-7R,8S-dihydroxy-9S,10R-epoxy-7,8,9,10-tetrahydrobenzo[a] at the hypoxanthine (guanine) phosphoribosyltransferase gene in repair-deficient Chinese hamster V-H1 cells," *Carcinogenesis*, vol. 20, no. 12, pp. 2279–2285, 1999.

[133] K.-Y. Seo, A. Nagalingam, M. Tiffany, and E. L. Loechler, "Mutagenesis studies with four stereoisomeric N^2-dG benzo[a]pyrene adducts in the identical 5′-CGC sequence used in NMR studies: $G \rightarrow T$ mutations dominate in each case," *Mutagenesis*, vol. 20, no. 6, pp. 441–448, 2005.

[134] R. E. Johnson, M. T. Washington, L. Haracska, S. Prakash, and L. Prakash, "Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions," *Nature*, vol. 406, no. 6799, pp. 1015–1019, 2000.

[135] F. Yuan, Y. Zhang, D. K. Rajpal et al., "Specificity of DNA lesion bypass by the yeast DNA polymerase η," *Journal of Biological Chemistry*, vol. 275, no. 11, pp. 8233–8239, 2000.

[136] S. Chandani and E. L. Loechler, "Y-family DNA polymerases may use two different dNTP shapes for insertion: a hypothesis and its implications," *Journal of Molecular Graphics and Modelling*, vol. 27, no. 7, pp. 759–769, 2009.

[137] S. Chandani and E. L. Loechler, "Molecular modeling benzo[a]pyrene N^2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4," *Journal of Molecular Graphics and Modelling*, vol. 25, no. 5, pp. 658–670, 2007.

[138] K. Y. Seo, J. Yin, P. Donthamsetti, S. Chandani, C. H. Lee, and E. L. Loechler, "Amino acid architecture that influences dNTP insertion efficiency in Y-family DNA polymerase V of E. coli," *Journal of Molecular Biology*, vol. 392, no. 2, pp. 270–282, 2009.

[139] F. Wang and W. Yang, "Structural insight into translesion synthesis by DNA Pol II," *Cell*, vol. 139, no. 7, pp. 1279–1289, 2009.

[140] M. Hogg, S. S. Wallace, and S. Doublié, "Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site," *EMBO Journal*, vol. 23, no. 7, pp. 1483–1493, 2004.