Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 87–102
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

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

Chiu Hong Lee [a], Sushil Chandani [b], Edward L. Loechler [b,*]

[a] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[b] *Biology Department, Boston University, MA 02215, USA*

Received 11 July 2005; received in revised form 21 October 2005; accepted 21 October 2005
Available online 4 January 2006

Abstract

Y-family DNA polymerases (DNAPs) are a superfamily of evolutionarily related proteins that exist in cells to bypass DNA damage caused by both radiation and chemicals. Cells have multiple Y-family DNAPs, presumably to conduct translesion synthesis (TLS) on DNA lesions of varying structure and conformation. The potent, ubiquitous environmental mutagen/carcinogen benzo[*a*]pyrene (B[*a*]P) induces all classes of mutations with G → T base substitutions predominating. We recently showed that a G → T mutagenesis pathway for the major adduct of B[*a*]P ([+ta]-B[*a*]P-N^2-dG) in *Escherichia coli* depends on Y-family member DNAP V. Since no X-ray crystal study for DNAP V has been reported, no structure is available to help in understanding the structural basis for dATP insertion associated with G → T mutations from [+ta]-B[*a*]P-N^2-dG. Herein, we do homology modeling to construct a model for UmuC, which is the polymerase subunit of DNAP V. The sequences of eight Y-family DNAPs were aligned based on the positioning of conserved amino acids and an analysis of conserved predicted secondary structure, as well as insights gained from published X-ray structures of five Y-family members. Starting coordinates for UmuC were generated from the backbone coordinates for the Y-family polymerase Dpo4 for reasons discussed, and were refined using molecular dynamics with CHARMM 27. A survey of the literature revealed that *E. coli* DNAP V and human DNAP η show a similar pattern of dNTP insertion opposite a variety of DNA lesions. Furthermore, *E. coli* DNAP IV and human DNAP κ show a similar dNTP insertional pattern with these same DNA lesions, although the insertional pattern for DNAP IV/κ differs from the pattern for DNAPs V/η. These comparisons prompted us to construct and refine models for *E. coli* DNAP IV and human DNAPs η and κ as well. The dNTP/template binding pocket of all four DNAPs was inspected, focusing on the array of seven amino acids that contact the base of the incoming dNTP, as well as the template base. DNAPs V and η show similarities in this array, and DNAPs IV and κ also show similarities, although the arrays are different for the two pairs of DNAPs. Thus, there is a correlation between structural similarities and insertional similarities for the pairs DNAPs V/η and DNAPs IV/κ. Although the significance of this correlation remains to be elucidated, these observations point the way for future experimental studies.

Keywords: Cancer; Genotoxins; Mutations; Adducts; Lesion; DNA polymerase; Pol IV; Pol V; Pol eta; Pol kappa

Abbreviations: aa, amino acids; (+)-*anti*-B[*a*]PDE, 7*R*,8*S*-dihydroxy-9*S*,10*R*-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene; (+)-*anti*-dB[*a*,j]ADE, the corresponding diol epoxide derived from dibenz[*a*,j]anthracene; AP sitesapurinic/apyrimidinic sites; B[*a*]P, benzo[*a*]pyrene; CPD, cyclopyrimidine dimer, a UV-photoproduct, characterized by the formation of a cyclobutane ring between the 5,6-positions of adjacent pyrimidines; DNAP, DNA polymerase; η, eta; ι, iota; κ, kappa; PAH, polycyclic aromatic hydrocarbon; (6-4) photoproduct, a UV-photoproduct, with a 6-4-pyrimidine-pyrimidone structure formed at adjacent pyrimidines; [+ta]-B[*a*]P-N^6-dA, a minor adduct of (+)-*anti*-B[*a*]PDE formed by *trans* addition of N^2-dA to C10 of (+)-*anti*-B[*a*]PDE; [+ta]-B[*a*]P-N^2-dG, the major adduct of (+)-*anti*-B[*a*]PDE, formed by *trans* addition of N^2-dG to (+)-*anti*-B[*a*]PDE; TLS, translesion synthesis: the insertion of a base opposite a DNA adduct, as well as subsequent elongation

* Corresponding author. Tel.: +1 617 353 9259; fax: +1 617 353 6340.
E-mail address: loechler@bu.edu (E.L. Loechler).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2005.10.009

1. Introduction

Recently, it has become apparent that cells possess an unexpectedly large number of DNA polymerases, e.g. human cells have at least fifteen, while *Escherichia coli* has at least five [1–4]. The cellular significance of some of these DNA polymerases (DNAPs) can be understood by noting that DNA is constantly subjected to insult by chemicals and radiation, and most lesions that are not removed by DNA repair block replicative DNA polymerases. To avoid such lethal blockage, cells possess a class of so-called ''lesion-bypass DNA polymerases'', which conduct translesion DNA synthesis (TLS [1–14]). Many of the DNAPs capable of TLS belong to the Y-family, which is a newly recognized superfamily of evolutionarily related DNAPs [1–14]. Human cells have at least four Y-family members, DNAPs η, ι and κ and REV1, while *E. coli* has at least two, DNAPs IV and V. Cells are believed to have multiple lesion-bypass DNA polymerases because of the structural and conformational diversity of DNA lesions. For example, while DNAP η bypasses cyclopyrimidine dimers (CPDs) accurately [15–20], it is DNAP κ that bypasses benzo[*a*]pyrene adducts accurately [21–26].

Y-family DNA polymerases range in size from ∼350 amino acids (aa) to over 1000aa, and share five conserved motifs (I–V) in their N-terminal ∼200 amino acids, which includes all of the residues in the template/dNTP binding pocket and polymerase catalytic active site (Fig. 1) [6]. They also share a number of important structural features, as revealed in X-ray crystallographic studies of the archael DNA polymerases Ddh [27] and Dpo4 [28–32], yeast DNAP η [33] and human DNAPs ι [34] and κ [35]. As with all DNA polymerases, Y-family members have thumb, palm and fingers domains. However, their fingers and thumb are relatively small or ''stubby'', resulting in greater solvent accessibility to the template/dNTP-binding pocket [12], which is more spacious, presumably to accommodate the bypass of bulky and/or deformed DNA lesions. Furthermore, Y-family polymerases grip DNA with an additional domain, which has been called the ''wrist'' [27], the ''polymerase associated domain (PAD)'' [33] or the ''little finger domain'' [28]. Y-family DNAPs appear to be in the ''closed'', replication-competent conformation normally, and may not use the induced-fit mechanism (with an accompanying conformational change) for dNTP discrimination as do replicative DNA polymerases [12,27,36,37].

We showed that the *E. coli* Y-family DNAP V is involved in a G → T mutational pathway for [+ta]-B[*a*]P-N^2-dG [38], which is the major adduct formed from (+)-*anti*-B[*a*]PDE, a metabolically reactive form of benzo[*a*]pyrene (B[*a*]P). B[*a*]P is a potent mutagen and carcinogen, and is an example of a polycyclic aromatic hydrocarbon (PAH), which is a class of substances produced by incomplete combustion that are found ubiquitously in the environment [39–46]. Most carcinogens are active by causing mutations, and it is generally believed that mutations induced by PAHs in general, and B[*a*]P in particular, can cause mutations relevant to cancer causation (representative references: 47–53), and are important in human cancer (e.g. 54 and references therein).

B[*a*]P mutagenesis has been extensively studied [55–59], and mutational spectra with the relevant metabolite (+)-*anti*-B[*a*]PDE has been determined in *E. coli* [60,61] and in mammalian (CHO) cells [62,63 and references therein], where base substitutions dominate, but insertions and deletions are also significant. The type of mutation induced by (+)-*anti*-B[*a*]PDE depends on DNA sequence context [60,61], which is also observed with the major adduct [+ta]-B[*a*]P-N^2-dG, with the most dramatic examples being observed in *E. coli*: G → T mutations dominated in a 5′-T<u>G</u>C-3′ sequence (>95% [64]) and G → A mutations dominated in a 5′-A<u>G</u>A-3′ sequence (∼95% [65]), while −1 frameshift mutations dominated in a 5′-GG<u>G</u>A sequence (∼70% [66]).

These results raise the question: by what mechanism(s) can a single adduct induce multiple kinds of mutations? This question is best addressed with examples from studies in *E. coli* with AAF-C8-dG, which is the major adduct of *N*-2-acetylaminofluorene [67–70]. AAF-C8-dG in a 5′-CGC<u>G</u> sequence either can cause a −2 frameshift mutation in a process dependent on DNA polymerase II (DNAP II), or can result in no mutation in a process dependent on DNAP V [67–70]. The current model is that AAF-C8-dG at a replication fork can exist in two different conformations: one conformation in which the adducted dG moiety forms a −2 slipped intermediate that includes a primer terminus that DNAP II can elongate or a second conformation in which the adducted dG is properly positioned to form a base pair with dC, as catalyzed by DNAP V. The simplest interpretation of these findings is that the pathway of lesion-bypass (be it error-free or mutagenic) for a single adduct depends on its conformation, which in turn dictates what DNAP is involved in the TLS event. Thus, the lesion-bypass event is most likely defined by both adduct conformation and the DNAP doing the bypass, and both factors must be understood in order to understand mutagenic mechanism.

In vitro studies with purified lesion-bypass DNA polymerases have revealed other principles. Different DNA polymerases can respond differently to a single DNA adduct as illustrated with B[*a*]P-N^2-dG adducts, where (for example) bypass of [+ta]-B[*a*]P-N^2-dG principally involved dCTP insertion with human polymerase κ [21–26], but dATP insertion with human polymerase η [21,71–73]. We showed that DNAP V plays a role in a G → T pathway in *E. coli* [38], and consistent with this finding, purified DNAP V was shown to preferentially insert dATP opposite [+ta]-B[*a*]P-N^2-dG in a 5′-C<u>G</u>T sequence context, while dATP and the correct nucleotide dCTP were inserted approximately equally in a 5′-G<u>G</u>T sequence context [74].

We wish to understand the structural basis for dATP insertion by DNAP V in order to understand the G → T mutational pathway with [+ta]-B[*a*]P-N^2-dG. No X-ray crystal structure exists for DNAP V, and it is unlikely to be forthcoming in the near term given that DNAP V remains difficult to purify and has complex properties [84,85 and references therein]. Herein, we describe homology modeling of UmuC, which is the polymerase subunit of DNAP V. In studying literature findings about TLS for six DNA adducts/lesions, we noted parallels between *E. coli* DNAP V (UmuC) and human DNAP η, as well

Fig. 1. Amino acid alignment of three Y-family DNAPs whose structure is unknown (hDNAPs η, *E. coli* DNAP IV and UmuC of DNAP V) with five DNAPs whose structure is known (Dpo4, Dbh, yDNAP η, hDNAP ι and hDNAP η). α Helices (blue), β sheets (yellow) and turns (green) are shown. Letters A–M above the alignment indicate conserved α helices, and numbers 1–12 indicate conserved β sheets in Dpo4, UmuC and DNAP IV; hDNAPs η, yDNAP η, hDNAP ι and hDNAP η have slightly different assignments, because of their extra regions. The ''linker'' connects the palm and little finger domains. Y-family DNAPs have five conserved motifs, which are indicated by the horizontal bars (in Dpo4, I: ∼3–30aa, II: ∼41–56aa, III: ∼78–110aa, IV: ∼123–161aa and V: 173–206aa). Backbone coordinates for (e.g.) UmuC derived from Dpo4 are indicated by blue letters, while green letters indicate coordinates generated by the Loop Generation function in the homology module of InsightII. Boxes indicate conserved amino acids (five of eight) or other notable amino acid similarities.

as between DNAP IV and human DNAP κ, so we have constructed models for each of these DNAPs. It is our goal to use observed structural differences and similarities between these DNAPs as the basis for experiments to probe the mechanistic basis for their differences in insertional patterns.

2. Methodology

Graphics were visualized using InsightII (Accelrys Inc. [75]) on a SGI O2. Molecular modeling and molecular dynamics were done with CHARMM 27 [76–78] and calculations were done either on an SGI Origin200 or one of Boston University's IBM Power4 systems (p690 or p655). The coordinates for Dpo4 (1JX4, 1.70 Å resolution [28]) are the basis for the initial structures for UmuC and DNAP IV. The coordinates for Dpo4 and yDNAP η (1JIH, 2.25 Å resolution [33]) are the basis for building hDNAP η. The coordinates for hDNAP κ (1T94, 2.40 Å resolution [35]) and yDNAP η are the basis for building hDNAP κ, although the positioning of the little fingers domain with respect to the thumb/palm/fingers domains was accomplished by superimposition on Dpo4 (see below).

We describe our protocol for UmuC as an example, before describing DNAP IV, hDNAP η and hDNAP κ. Our alignment (Fig. 1) began with the alignment of Woodgate and co-workers [6], which we refined by inspecting the known X-ray structures for structural elements that should be preserved. Secondary structural predictions algorithms were also used to confirm which regions were likely to be α versus β versus loops, as well as the positioning of the likely borders. Both Chou-Fasman [108] and GOR II [109] were used from the homology module of InsightII. Finally, following MD runs Profiles-3D was used to evaluate the final structures, and regions with low Profiles-3D scores were evaluated to assess whether realignments were needed; several examples of this iterative alignment procedure are described at the end of this section. Pairwise comparisons between Y-family DNAPs show that amino acid sequences are typically ∼25% identical/∼45% similar (Table 1, which is based on the alignment in Fig. 1, as analyzed by CLUSTAL W [110]). Based on the alignment in Fig. 1, initial, input coordinates were assigned for UmuC. Where alignments agreed between Dpo4 and UmuC, the positions (coordinates) of the backbone atoms in the amino acid chain were made the same in the initial UmuC structure (Fig. 1: blue letters). Coordinates for the non-conserved side chains were assigned using the Coordinate Assigning function in the homology module of Insight II. In places where UmuC had more or fewer amino acids than Dpo4 (always in loops), new loops for UmuC were generated using the Loop Generation function in the homology module in Insight II. Missing hydrogens were added using Biopolymers in Insight II. Neither water nor counter ions were included because in this study we were primarily interested in the positioning of amino acids around the adenine in the dATP, where there is no room for water. In the future, we will use the output structures generated as described below as starting structures for more extensive molecular modeling with water and counter ions.

After constructing the starting protein coordinates, UmuC was superimposed on the coordinates for Dpo4, and the DNA coordinates from Dpo4 were added. The Dpo4/1JX4 X-ray structure has ddADP in its active site, and the following was done to replace it with dATP. Three lesion-containing Dpo4 structures have dATP, but none of these fit optimally into the Dpo4/1JX4 active site, as the B-peptide of Dpo4 with a TT-CPD [37d: 1RYS] had *syn*-adenine, while the B-peptide of Dpo4 with a B[*a*]P adduct [37e: 1S0M] and one structure containing an apurinic site (AP site) [37f: 1N56] did not fit well, apparently due to slight distortions in Dpo4 induced by the lesions. The approach that worked best involved the docking into Dpo4/1JX4 of the coordinates for dNTP from the T7 DNAP X-ray structure, along with its two catalytic magnesiums and their two crucial aspartate ligands, which are designated D475 and D654 in the T7 X-ray coordinate file (1T7P [79]). D475 and D654 of T7 superimposed closely on D7 and D105 of Dpo4, respectively (atoms <1 Å apart). (The D7 and D105 equivalent positions are conserved in all Y-family DNAPs (Fig. 1).) It is important to note that the shapes of dATP from Dpo4/1RYS, Dpo4/1S0M and Dpo4/1N56 are very similar to T7 DNAP, and inclusion of any of them would undoubtedly have made little difference in the final structures. The docked dATP (base constrained) was minimized (Conjugate Gradient, tolerance gradient <0.01) with DNA and Dpo4 constrained, and this Dpo4 structure was used as the basis for modeling the other structures.

The starting coordinates for UmuC were refined using a CHARMM 27 all atom forcefield as follows. First, the protein backbone, the incoming dATP (except for the β and γ

Table 1
Amino acid identities and similarities in pairwise comparisons for Y-family DNAPs

	Dpo4	DBH	hDNAP ι	yDNAP η	hDNAP η	hDNAP κ	UmuC (V)	DNAP IV
Dpo4	–	54/79	27/52	19/28	24/50	28/49	29/47	30/52
DBH	–	–	27/47	20/45	21/45	24/47	23/43	28/52
hDNAP ι	–	–	–	22/47	28/50	25/44	21/37	29/53
yDNAP η	–	–	–	–	24/45	18/41	16/31	20/42
hDNAP η	–	–	–	–	–	24/41	20/40	30/50
hDNAP κ	–	–	–	–	–	–	21/35	36/57
UmuC (V)	–	–	–	–	–	–	–	19/42
DNAP IV	–	–	–	–	–	–	–	–

Pairwise comparisons made by CLUSTAL W analysis (108) based on the alignments in Fig. 1 with the exception that that any extra unaligned sequences (gaps) were ignored in the scoring. The first value is amino acid identity and the second is amino acid similarity.

phosphates), and the DNA (except for the single stranded region) were constrained. Steepest Descent (10 K steps) and Conjugate Gradient (20 K steps) minimization were performed. Thereafter, the constraints on the amino acid backbone were released, and Conjugate Gradient minimization (20 K) was repeated. Each structure was heated to 300 K (1 ps) and equilibrated (1 ps), after which molecular dynamics was conducted at 300 K for 500 ps, at which point the energy was no longer declining as a function of time. Structures were then cooled to 0 K over 6 ps, and finally each structure was energy minimized (Conjugate Gradient for 20 K steps).

E. coli DNAP IV was built similarly to UmuC.

Even though a crystal structure exists for yeast DNAP η, we have based our human DNAP η model principally on the Dpo4 coordinates for five reasons. (1) In spite of the names, hDNAP η overall is no more related to yDNAP η (24% identity/45% similarity) than it is to Dpo4 (24% identity/50% similarity) (Table 1). (2) Importantly, the Dpo4 X-ray structure has DNA, while yDNAP η does not. (3) It appears that the X-ray structure for yDNAP η must undergo a significant conformational change in order to bind DNA. The little finger domain must move ∼20 Å toward the thumb domain, while the thumb domain must move ∼8 Å toward the palm domain (data not shown). These conformational changes were noted by others who built a model of yeast DNAP η with DNA [29]. (4) The X-ray structures for Dpo4 and yDNAP η include 341aa and 509aa, respectively, and yet, the alignment of these portions of the proteins corresponds almost exactly (Fig. 1) because yDNAP η has approximately six extra loops/regions/domains not present in Dpo4. Three regions present in yeast DNAP η but absent in Dpo4 are large (70, 34 and 14 amino acids), while three are smaller (six amino acids each). In contrast, human DNAP η (∼405aa) is actually closer in size to Dpo4 than to yDNAP η and the difference in their sizes is virtually entirely due to a single extra region (aa120–189) present in hDNAP η that is absent in Dpo4. Thus, hDNAP η has one region not present in Dpo4, while yDNAP η has three large regions (six total regions) not present in Dpo4, and yDNAP η has two (or five) regions not present in hDNAP η. (5) The ''roof-aa'', which stacks on the base of the incoming dATP, is particularly important for our subsequent analysis (see Section 4), and is A44 in Dpo4. On the N-terminal side of the roof-aa, Dpo4 has a larger loop by 7-aa compared to yDNAP η (Fig. 2C). hDNAP η also has a larger loop by 5-aa, making it closer in size to Dpo4 than to yDNAP η. For these five reasons, we constructed our model of human DNAP η from Dpo4 rather than from yeast DNAP η.

Initial coordinates for the extra region of human DNAP η (aa120–189) not in Dpo4 was built from the coordinates for the analogous region of yeast DNAP η (aa176–237) and inserted in place of aa116–117 of Dpo4. As discussed in Section 3, the X-ray structures for yDNAP η and Dpo4 show that this insertion point makes sense, as the backbone coordinates for yDNAP η (aa146–256) can be superimposed almost precisely on the analogous region of Dpo4 (aa96–136) (Fig. 2C). Thus, hDNAP η was built based on the structural fusion protein [Dpo4(1–115)/yDNAP η(176–237)/Dpo4(118–341)].

Recently, a structure for hDNAP κ with no DNA was published [35]. The structure, however, has several significant distortions, making it unsuitable for starting coordinates in our modeling study. Most notably, the little finger domain is ''broken/dislocated'', as if it were pointing out from the back of the hand. Much of the thumb, palm and fingers domains of hDNAP κ superimposed reasonably well on Dpo4, and this allowed us to re-position the hDNAP κ little finger using the relative positioning of the Dpo4 little finger as a guide. ''Linker'' is the term used to describe the amino acids that connect the palm domain to the little finger domain. The linker amino acids are missing in the hDNAP κ structure, and were built using the Loop Generation function in the homology module in Insight II with the Dpo4 linker as a guide. In the hDNAP κ X-ray structure, amino acids 211–288, which are near the top of the palm, appear to be significantly mis-positioned, and stick out of the back of the hand, with some amino acids missing. After considerable analysis, we concluded that this region is not only mis-positioned, but also likely mis-folded, so it could not be simply re-positioned intact. Thus, the aa211–288 region was re-constructed using yDNAP η as a guide in a manner analogous to the approach described above for hDNAP η. Finally, several other stretches of amino acids, principally in the little finger domain, were missing and were added using the Loop Generation function in the homology module in Insight II. For the thumb/palm/fingers domains, we used the A-isomer from the X-ray coordinates of the hDNAP κ dimer, since the A-isomer appeared more complete and better structured than the B-isomer of the dimer. In contrast, we used the B-isomer coordinates for the little finger domain, since it had fewer gaps, and appeared more intact and similar in structure to other little finger domains. We note, however, that the coordinates for A-isomer and the B-isomer are very similar.

The MD refined output structures were evaluated by Profiles-3D analysis, which scores each amino acid based on 18 environmental categories, as derived from composite information about amino acid positioning in proteins of known structure [80–83]. Positive values for a particular amino acid are usually viewed as acceptable. Two examples of how we used Profiles-3D to evaluate and improve input starting structures follow.

Example #1 is enumerated in four steps. Step 1: our earliest input coordinates for UmuC were based on aligning UmuC with Dpo4 (Fig. 1, call it: ''alignment #1''), and, following molecular dynamics, the output structure had better Profiles-3D scores for virtually all UmuC residues, except aa300–316, where scores were very unfavorable (in the range of −0.4 to −0.6). We thought this might be due to having Proline315 embedded in an α helix. (Dpo4 has no proline at the equivalent position.) Step 2: we unfolded a small portion of the helix around P315 in the input coordinates and conducted another MD run, but this gave another unacceptable output structure as assessed by Profiles-3D analysis (data not shown). Step 3: additional secondary structural predictions suggested that

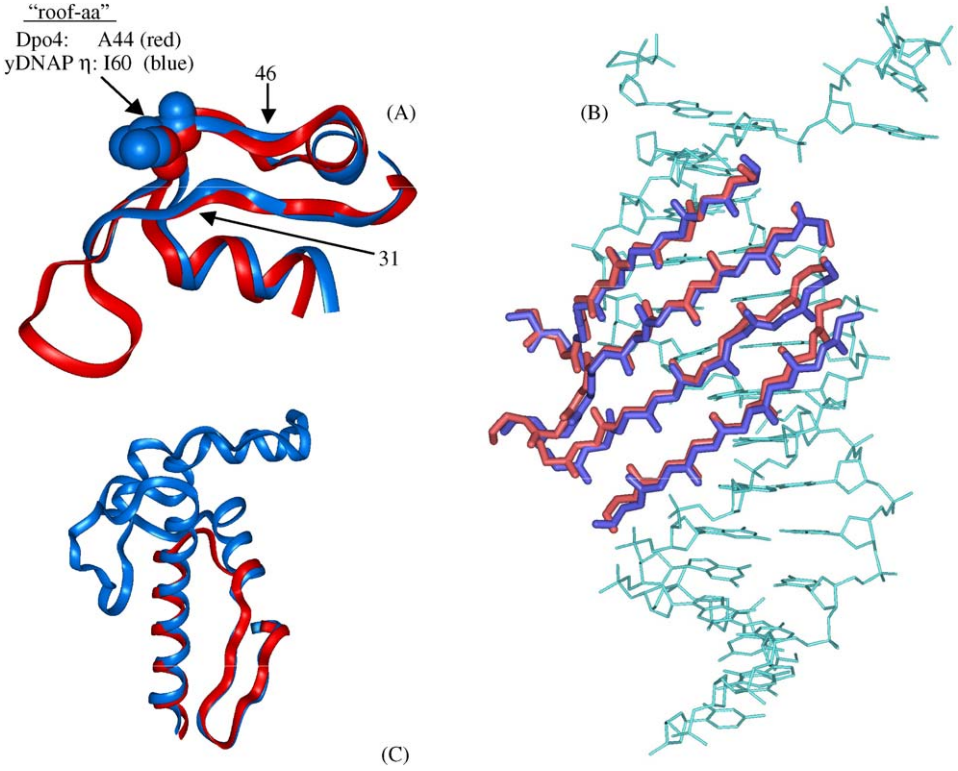

Fig. 2. Examples highlighting the conservation of structure for Y-family DNAPs as revealed in their X-ray coordinates. (A) A ribbon diagram comparison of Dpo4 (red: aa24–69) with yeast DNAP Eta (blue: aa46–85). Roof-aa are shown as spheres (A44/red for Dpo4; I60/blue for yeast DNAP η). No amino acid identity exists for Dpo4 and yDNAP η between aa-31 and aa-46 (based on the Dpo4 numbering system; Fig. 1). The view is approximately up the DNA helix axis, as if looking up at the ceiling of the active site. (B) Comparison of four parallel β strands in the little finger domains of Dpo4 (blue) and hDNAP ι (red). These β strands fit into the major groove of DNA (turquoise). Only the backbone amino acids are shown. (C) A ribbon diagram comparison of Dpo4 region aa96–136 (red) with yDNAP η aa146–256 (blue).

residues 307–313 in UmuC might in fact be β, and when the entire aa308–327 region was converted to a β sheet and β turn (call this: ''alignment #2''), the resulting output UmuC structure (following MD) had good Profiles-3D scores. However, the structure had a problem. β Sheets require fewer amino acids to traverse a fixed distance in comparison to α helices, and consequently only 332 amino acids from UmuC were required to match the amino acids in the Dpo4 crystal structure. Importantly, in Dpo4, R336 and K339 interact with phosphates in the DNA backbone, and the UmuC output structure based on the original alignment #1 had conserved amino acids R339 and K342 also interacting with these phosphates. In alignment #2, however, equivalent basic amino acids were not present in the UmuC output structure, and following molecular dynamics, the UmuC structure looked inferior as a result. Step 4: returning to alignment #1, we thought that the problem in the aa300–316 region might be due to a bad turn/loop at aa301–307. A new loop was constructed (rather than accepting the backbone coordinates based on aa301–307 in Dpo4). Following molecular dynamics, the output UmuC structure had improved Profiles-3D scores (all positive), appeared reasonable, and importantly retained the R339/K342-phosphate interactions. Incidentally, the aa301–307 region is >25 Å from the active site, is on the backside of UmuC in the little finger domain, and is likely to be involved in an interaction with other proteins (e.g. UmuD', which is the regulatory subunit of DNAP V).

Example #2 involves aa63–72 in UmuC, where two alignments were tried. We prefer alignment #1, which is shown in Fig. 1, where the loop between helix-C/strand-4 has two additional amino acids compared to other DNAPs. The advantage of alignment #1 is that it allows the final UmuC structure to retain a tertiary interaction (V68/L22) that is present in Dpo4 (V72/L23) and most of the other DNAPs (Fig. 1). In alignment #2, which is not shown in Fig. 1 and we do not prefer, R63 is located two aa-positions to the right of where it is alignment #1 (Fig. 1), such that R63 aligns with P69 in Dpo4. The advantage of alignment #2 is that it places the extra amino acids present in UmuC compared to Dpo4 into a loop between strand-4 and helix-D, where three of the other Y-family DNAPs are known to have extra amino acids compared to Dpo4 based on X-ray structures as shown in the alignments in Fig. 1. One disadvantage of alignment #2 is that G66 must therefore interact with L22 in UmuC, with the consequences being that no favorable hydrophobic interaction is found in the final UmuC structure. For aa63–72 of UmuC, the Profiles-3D scores clustered around a higher value in alignment #1 (∼0.2) compared to alignment #2 (∼0.0). Thus, we believe that alignment #1 is more likely to be correct based on the Profiles-3D scores and the retention of the V68/L22 interaction.

Many other alignments were also tried and structures generated, but they were not favored based on these same kinds of assessments of Profiles-3D scores and structural considerations.

3. Results

When we initiated this molecular modeling study, only the X-ray crystal structures for the Y-family DNAPs Dpo4, Dbh and yDNAP η were available in the RCSB Protein Data Bank. Our models were built principally based on the coordinates for Dpo4 (1JX4, 1.7 Å resolution [28]), because at the time it was the only DNAP that contained DNA, and evidence suggests that conformational changes occur in Y-family DNAPs upon DNA binding (see Section 2). Since initiating our work, Dpo4 structures have been published with lesion-containing DNA (i.e. TT-CPD [29: 1RYR and 1RYS], a B[*a*]P adduct [30: 1S0M], an AP site [31: 1N48, 1N56, 1S0N, 1S0O and 1S10] and a G:T mismatch [32: 1S97]). However, we still wished to begin our efforts using lesion-free DNA, and, thus, Dpo4/1JX4 have been used as the primary basis for our homology modeling. Recently, a structure for DNAP ι with DNA [34] emerged, and we intend to use it as a model in the future.

While the motivation for the study herein was our desire to understand dATP insertion opposite [+ta]-B[*a*]P-N^2-dG, the X-ray structure of Dpo4 with a B[*a*]P adduct (1S0M) was not used for two reasons. (1) The B[*a*]P moiety in Dpo4/1S0M is in the duplex, double stranded region of the DNA (i.e. in the [n + 1] position) and not in the active site, so that nothing is revealed about how a dNTP might be incorporated opposite a B[*a*]P adduct. (2) The structure contains a *cis*-opened BP-N^6-dA adduct derived from (+)-*anti*-B[*a*]PDE, which is a minor adenine adduct that only superficially resembles [+ta]-B[*a*]P-N^2-dG.

The primary amino acid sequences were aligned for eight Y-family DNAPs by Yang and co-workers [6], and we took their alignment and refined it further to give the results in Fig. 1 (Section 2), which shows the N-terminal portions that are homologous to the first ∼350aa of Dpo4. These eight DNAPs include the five for which X-ray structures have been published, as well as the four we wished to model. Conserved amino acids (Fig. 1, five of eight amino acids in agreement boxed) primarily appear in the N-terminal regions of the protein, which contains the active site, the thumb/palm/fingers domains, and the five conserved Y-family motifs (i.e. in Dpo4, I: ∼3–30aa, II: ∼41–56aa, III: ∼78–110aa, IV: ∼123–161aa and V: 173–206aa). Alignment was also evaluated by locating boundaries of conserved structural elements (i.e. α helices, β sheets and turns) in the known X-ray structures for Y-family DNAPs and using secondary structural prediction algorithms (e.g. Chou–Fasman rules) to locate similar regions and boundaries (Section 2). Fig. 1 shows the positioning of α helices (blue), β sheets (yellow) and turns (green) in the known structures, and where they were assigned in our modeled structures. In places where gaps or extra amino acids are apparent, turns are virtually always observed or predicted. Thus, these conserved proteins seem to share a conserved core backbone region, and differ principally in that they have variable loops.

For our homology models to be reasonable, the alignments in Fig. 1 must be accurate. Pairwise comparisons between Y-family DNAPs show that amino acid sequences have on average ∼25% identity/∼45% similarity (Table 1). In viewing the published structures of Y-family DNAPs, however, it appears that, although amino acid identity is often modest, structural similarity is strikingly high. Three examples illustrate this point.

Example #1: regions 7–30 and 47–69 show considerable homology for the Y-family DNAPs as aligned in Fig. 1 (numbering based on Dpo4). In contrast, homology is low for the 31–46 region. However, even though (e.g.) Dpo4 and yDNAP η have no identical amino acids, their X-ray coordinates superimpose very closely (Fig. 2A), in spite of the fact that Dpo4 has a larger loop.

The alignment in this region is particularly important, because of the ''roof-aa'', which is the amino acid that sits on top of the base of the incoming dNTP in the active site. The roof-aa with Dpo4 (A44) and yDNAP η (I60) are virtually superimposeable in Fig. 2A.

To reinforce the argument that the alignment around the roof-aa is sensible, we note that there appear to be two patterns of conservation for the roof-aa and the next aa (Fig. 1): a roof-isoleucine followed by alanine (e.g. I38/A39 with UmuC) or a roof-aa that is less bulky than isoleucine followed by threonine (e.g. S42/T43 with DNAP IV). Furthermore, an amino acid with a bulky alkyl-side chain is present before the roof-aa (i.e. L, I or V).

Example #2: homology is low when comparing the little finger domains (Fig. 1, aa244–341 with Dpo4), and yet the X-ray coordinates reveal striking structural similarities. For example, four structurally conserved, β strands fit into the major groove of DNA, as illustrated in Fig. 2B for Dpo4 (red) and hDNAP κ (blue), which show R.M.S.D. = 0.83 Å for the backbone atoms. This is particularly striking, because it is in spite of the fact that the little finger domain of hDNAP κ is significantly out of position as revealed in the X-ray structure (discussed in Section 2).

Example #3: the eukaryotic Y-family DNAPs have an extra domain (e.g. aa176–237 in yeast DNAP η) not present in the prokaryotic DNAPs (Fig. 1). A comparison of the X-ray structures for yDNAP η and Dpo4 show that the backbone coordinates for yDNAP η (aa146–256) can be superimposed almost precisely on the analogous region of Dpo4 (aa96–136) (Fig. 2C), in such a way that it looks like the entire yDNAP η region (aa146–256) could have evolved by removing (aa116–117) from Dpo4 and substituting yDNAP η (aa176–237). We note that this extra domain is never closer than ∼25 Å from the active site in yDNAP η, suggesting that it is involved in structure/regulation and not directly involved in (e.g.) dNTP selection.

These and other comparisons suggest that the overall structures of Y-family DNAPs are strikingly well conserved and that the alignments in Fig. 1 are likely to be largely correct. This is especially true for regions with greater consensus, notably sequences that are in alignment with the first ∼200 amino acids of Dpo4 (Fig. 1), which include the catalytic site and the template/dNTP binding pocket (see below). The little finger domains (above ∼aa240 in Dpo4) have a paucity of conserved amino acids, making this region of greatest uncertainty. We note, however, that we aligned the little finger domain of

hDNAP κ prior to the publication of its X-ray structure [35], and our alignment of β strands 9–12 and α helices L and M were exact.

Based on the alignments in Fig. 1, initial, input coordinates were assigned. Where alignments agreed between Dpo4 and (e.g.) *E. coli* UmuC, the positions (coordinates) of the backbone atoms in the amino acid chain were made the same in the initial UmuC structure (Fig. 1: blue letters). In regions where UmuC had more (or fewer) amino acids than Dpo4, new larger (or smaller) loops were generated as described in Section 2 (Fig. 1: green letters). *E. coli* DNAP IV was built similarly to UmuC.

Even though a crystal structure exists for yeast DNAP η, we based our human DNAP η model principally on the Dpo4 coordinates for five reasons as discussed in Section 2. Importantly, in spite of the names, hDNAP η is more similar to Dpo4 than to yDNAP η based on a variety of criteria, and DNA is in the Dpo4 X-ray structure, but not in the yDNAP η structure. Recently, a structure for hDNAP κ without DNA was published [35], and, although it has numerous deformities (see Section 2), it could be used to build a starting structure (Section 2).

After constructing the starting protein coordinates for UmuC, DNAP IV, hDNAP η and hDNAP κ, each was superimposed on the coordinates for Dpo4, and the DNA coordinates from Dpo4 were added. The Dpo4/1JX4 X-ray structure has ddADP, which was replaced by dATP as described in Section 2. The starting coordinate structures for UmuC, DNAP IV, DNAP η and DNAP κ were each subsequently refined by a protocol that included a MD run at 300 K using a CHARMM 27 all atom forcefield (Section 2). Since the main focus of the work herein was to visualize the template/dNTP binding pocket, in particular to see the arrangement of amino acids around the base of the dNTP, water was not included because it would substantially slow down the calculations and there is no room for water in this pocket. The final structure for UmuC, which is the polymerase subunit of DNAP V, is shown in Fig. 3 as an example. (Coordinates are posted on our website.)

The MD-refined output structures were evaluated by Profiles-3D analysis, which scores each amino acid based on 18 environmental categories, as derived from composite information about amino acid positioning in proteins of known structure [80–83]. Two examples of how we used Profiles-3D to evaluate and iteratively improve alignments and structures are presented in Section 2. Following MD refinement, Profiles-3D scores for each DNAP were dominantly favorable, being

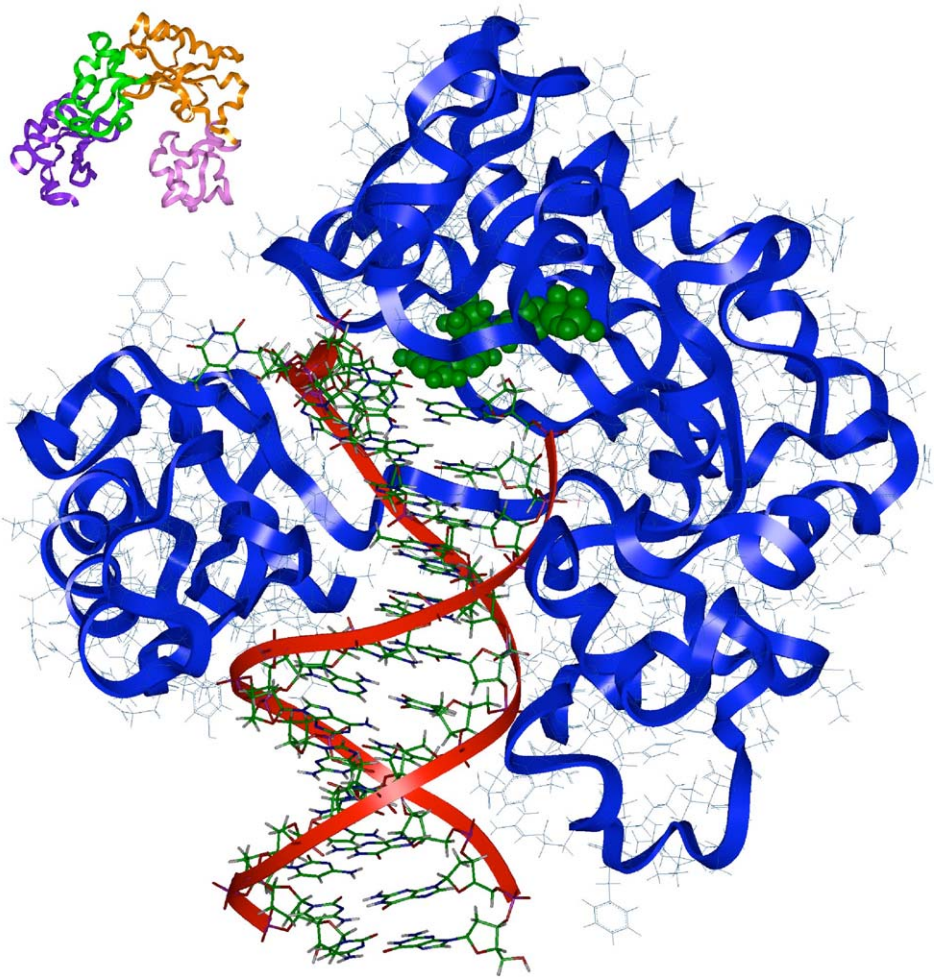

Fig. 3. Model of UmuC, showing the protein (blue ribbon), the DNA (red ribbon) and the dATP (green). *Insert*: top view of the UmuC model without DNA, showing the thumb (pink), palm (brown), fingers (green) and little finger (purple) domains. The right-handed nature of the structure is apparent.

especially good for DNAP κ (98.9%) and DNAP η (100%). Three DNAPs still had some negative values, although none lower than −0.15. The number of amino acids with negative Profiles-3D scores was as follows ([0.0 to −0.05]/[−0.05 to −0.1]/[−0.1 to −0.15]): UmuC (9/11/0, all in the little finger domain), DNAP IV (5/3/0), DNAP κ (2/0/4) and DNAP η (0/0/0).

4. Discussion

The investigation herein was prompted by our observation that DNAP V is involved in G → T mutations with [+ta]-B[*a*]P-N^2-dG in a 5′-TGT sequence [38], and by our desire to understand the structural basis for dATP insertion opposite this adduct to give a G → T mutation. Below we review the literature on dATP insertion by DNAP V (UmuC) opposite B[*a*]P adducts and other lesions and find an interesting pattern. This pattern prompted us to compare the biochemistry of UmuC with other Y-family DNAPs to assess whether other revealing patterns might emerge, and below we review the literature and report that DNAP V and hDNAP η appear to be functional orthologs, while DNAPs IV and κ appear to be functional orthologs based on their dNTP insertional patterns opposite DNA lesions (Table 2). (One caveat to acknowledge is that the similarities discussed below are not perfect, and that the comparisons involve findings from different laboratories using different sequence contexts for the DNA adducts/lesions, etc.) Finally, we consider our models of DNAPs IV, V η and κ to search for structural elements that might contribute to these bypass similarities.

4.1. DNAP V and its polymerase subunit UmuC

Purified DNAP V preferentially inserted dATP opposite [+ta]- and [−ta]-B[*a*]P-N^2-dG in a 5′-CGA sequence context [74]. While DNAP V preferentially inserted the correct base opposite [+ta]- and [−ta]-B[*a*]P-N^2-dG in a 5′-GGT, as well as opposite [+ta]- and [−ta]-B[*a*]P-N^6-dA in a 5′-TAG sequence, the most frequent misinsertion is dATP, which ranges between ∼8 and 40% in all four cases [74]. Purified DNAP V preferentially inserted dATP opposite both the 5′-T and 3′-T in a TT-CPD, and the 5′-T in a T(6-4)T [86]. dGTP and dATP (∼6:1) were preferentially inserted opposite the 3′-T in a T(6-4)T [86]; in a T(6-4)T, the 3′-T is a pyrimidone, whose structure looks like it should preferentially base pair with dG, although the G:T-pyrimidone pairing is strained as studied by NMR [87]. Finally, DNAP V preferred dATP insertion opposite an AP site, although dGTP is also inserted ([dATP/dGTP] = ∼2) [86].

DNAP V's behavior in cells is known. DNAP V is involved in error free bypass (dCTP insertion) with AAF-C8-dG in two sequences (5′-GGGA and 5′-GCGC) [88–91]. SOS-induction of *E. coli* led to DNAP V enhancement of −1 frameshift mutations with [+ta]-B[*a*]P-N^2-dG in a 5′-GGGA sequence [66], presumably via a Streisinger slippage bypass event in the run of dGs; the mechanism is thought to involve dCTP insertion, although not across from the adduct [89]. Finally, we showed that DNAP V is involved in G → T mutations (dATP insertion) opposite [−ta]-B[*a*]P-N^2-dG in a 5′-TGT sequence [38].

Thus, it appears that, opposite lesions, DNAP V either (i) preferentially inserts dATP or (ii) inserts the correct dNTP with the most frequent misinsertion being dATP. This analysis suggests that DNAP V has some kind of default dA insertional mode that was apparent with [+ta]- and [−ta]-B[*a*]P-N^2-dG, with [+ta]- and [−ta]-B[*a*]P-N^6-dA, with the 3′-T in a T(6-4)T, and with an AP site. (dATP insertion opposite TT-CPDs, and the 5′-T in a T(6-4)T could be either correct dNTP or default dATP insertion.)

The preferential insertion of dATP opposite several DNA lesions in *E. coli* led Dr. Bernard Strauss to coin the phrase the "A-rule" [92,93 and references therein]. The "A-rule" rationalizes UV light mutagenesis: C → T mutations predominate in 5′-PyC sequences, implying dATP insertion. UV mutagenesis depends on the genes *umuD/C*, which encode DNAP V, implying that DNAP V is required for UV mutagenic dATP insertion (discussed in Ref. [2]). Thus, dATP insertion by DNAP V seems likely to be the basis for *E. coli*'s A-rule. We speculate that UV light is probably the most frequently encountered form of DNA damage for which a lesion-bypass DNAP is important, and since thymine dimers (TT-CPDs) are the major UV lesion [94], a default dA insertion mode might help minimize UV mutagenesis.

What might be the basis for these two insertional modes? Regarding correct insertion, the Y-family member yDNAP η uses Watson–Crick hydrogen bonding between the incoming dNTP and the base being bypassed [36,37], which is not surprising and seems likely to be the case for Y-family DNAPs in general. We note that NMR studies have revealed that the dominant conformation for [+ta]-B[*a*]P-N^2-dG (which we call "BPmi5") shows more-or-less normal base pairing between the adduct-dG and its complementary dC [78 and references therein], and bypass of BPmi5 by a Y-family DNAP might then be expected to result in no mutation. We do not know the structural basis for UmuC's second mode, default dATP insertion. It might be either via some kind of mispairing mechanism (e.g. adduct-*syn*-G:A) or via a non-informational mechanism, whereby dATP is inserted when DNAP V cannot interpret the lesion.

Table 2
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 ≥ 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 insertion is indicated with references in the text. "n" indicates "no" or low activity; "A*" indicates bypass by an unusual mechanism (see text); "–" indicates data unavailable.

4.2. Human DNAP η resembles E. coli DNAP V (UmuC) in terms of dNTP insertional pattern opposite DNA adducts/lesions

Like DNAP V, hDNAP η also inserts dATP opposite both thymines in a TT-CPD [15,17,20,86], and preferentially incorporates dCTP opposite AAF-C8-dG, as does DNAP V both in vitro and in cells [88–91,95,96]. Both DNAPs V and η insert dGTP opposite the 3′-pyrimidone in a T(6-4)T [86,97]. hDNAP η inserts dATP opposite an AP site [73,95,97], as does DNAP V [86]. However, there are several subtle differences. DNAP V inserts dATP opposite the 5′-T of a T(6-4)T, while DNAP η is inactive [86,97]. DNAP V inserts dATP or the correct dNTP opposite B[*a*]P-N^2-dG adducts [38,74], while hDNAP η principally inserts dATP opposite [+ta]-B[*a*]P-N^2-dG in some sequences [21,72,73], or dATP and dGTP equally in others [71]. Thus, unlike DNAP V, there is no evidence currently that DNAP η has a correct dNTP insertional mode for B[*a*]P-N^2-dG adducts.

In summary, DNAP V seems to have two insertion modes: correct dNTP insertion and default dA insertion. hDNAP η also has a correct insertional mode (with some, though not all adducts), as does yeast DNAP η, which inserts dATP opposite a TT-CPD [16,18,19], dCTP opposite AAF-C8-dG [98] and dGTP opposite a CC-CPD [99]. hDNAP η seems to have a dATP insertion mode in some cases, although the preference is not as strong as DNAP V, as dATP/dGTP may both be added.

These comparisons (Table 2) suggest that dNTP insertion opposite lesions by hDNAP η is similar (although not identical) to DNAP V, suggesting that they can be thought of as functional orthologs. UmuC of DNAP V and hDNAP η show 20% identity/40% similarity, which is low for protein sequence comparisons between pairs of Y-family DNAPs (Table 1), so the functional similarity is presumably the result of subtler aspects of protein relatedness (see below).

4.3. Human DNAP κ resembles E. coli DNAP IV in terms of dNTP insertional pattern opposite DNA adducts/lesions

Human DNAP κ preferentially inserts dC opposite [+ta]-B[*a*]P-N^2-dG [21–26], which seems biologically relevant, given that DNAP κ was shown to be responsible for minimizing B[*a*]P-induced mutations in mammalian cells [25,26]. Purified DNAP IV preferentially inserts dCTP in two sequences (misinsertion in parenthesis): 5′-CGA (0.1%) and 5′-GGT (1.6%). For [−ta]-B[*a*]P-N^2-dG misinsertion is: 5′-CGA (1%) and 5′-GGT (5%) [74]. Both DNAPs IV and κ insert dCTP opposite AF-C8-dG [100], and both preferentially insert dTTP and dCTP (compared to dGTP and dATP) opposite AAF-C8-dG [100–103], although the ratio varies somewhat depending on sequence context. In Table 2, entries with ‘‘n’’ indicate where DNAPs IV [86] and κ [101,102] are known not to insert readily opposite a lesion. There is one exception in the similarity of lesion-bypass, although not without a caveat: DNAP IV does not bypass an AP site [86], while DNAP κ does [101,102], though the latter involves a slippage mechanism, requiring a specific sequence (5′-T), which was not tested in the case of DNAP IV.

Based on these comparisons (Table 2), DNAPs IV and κ appear to be functional orthologs, which is sensible, since hDNAP κ was originally identified as the closest human Y-family relative of *E. coli* DNAP IV [104–107], where DNAP IV and hDNAP κ have relatively high 36% identity/57% similarity (Table 1).

4.4. Charting the amino acids surrounding the ‘‘roof-amino acid’’ in DNAPs IV, V, η and κ

Thus, DNAPs V and η are similar in terms of lesion-bypass specificity, and DNAPs IV and κ are similar, while the DNAPs V/η set inserts differently than the DNAPs IV/κ set (Table 2). Undoubtedly, there are amino acid residues that define these similarities and differences. When understood in detail, this issue will probably prove to be complex and involve many facets of protein structure. For example, the protein region responsible for whether a particular DNA adduct/lesion can be accommodated by a particular DNAP (i.e. whether bypass is possible) may be different than the protein region that defines what dNTP is inserted in the bypass step.

Because of our studies on B[*a*]P adduct mutagenesis, we are most interested in the dATP default insertion mode noted for DNAP V, which seems most likely to be defined by amino acids that lie in the vicinity of the dNTP, and, thus, this region is considered next. Ultimately, our goal is to understand what might govern dNTP insertion with UmuC and other Y-family dNTPs.

In our model of UmuC (the polymerase subunit of DNAP V) the side chains of I38, M51 and G52 directly contact the adenine of the incoming dATP (Fig. 4A and B). In Dpo4, alanine-44 occupies the ‘‘roof-aa’’ position in that it stacks most directly on top of the adenine of dATP, and the equivalent position in UmuC is isoleucine-38 (purple, Fig. 4A and B). Methionine-51 (dark purple, Fig. 4A and B) also contacts the adenine of dATP, and we refer to it as ‘‘major groove 1 amino acid’’ (MaG1-aa), since it is oriented toward the developing major groove with respect to the roof-aa. We conducted multiple MD runs with the alignment in Fig. 1 and other alignments, and M51 adopted multiple positions. For example, the R-group of M51 can lie at the side of the adenine of dATP (Fig. 4A and B), or it can insert into the space above the adenine, where it either shared stacking with I38 (Fig. 4C and D) or dominated stacking, forcing I38 toward the developing minor groove (Fig. 4E and F). The α-carbon of G52 shows a small amount of contact with the adenine (light purple, Fig. 4A–F), and is oriented toward the developing major groove (MaG2-aa). C36 stacks on the template thymine (data not shown) and is referred to as the ‘‘template-aa’’.

The other amino acids that sit around the ‘‘roof-aa’’ can be analyzed in pairs. Two amino acids (Fig. 4G) contact the roof-aa and sit above the deoxyribose of the dATP, namely V29 (dR1-aa, red) and A39 (dR2-aa, dark red). Two amino acids (Fig. 4H) contact the roof-aa and lie in what will become the minor groove, namely S31 (MiG1-aa, yellow) and S72 (MiG2-aa, blue).

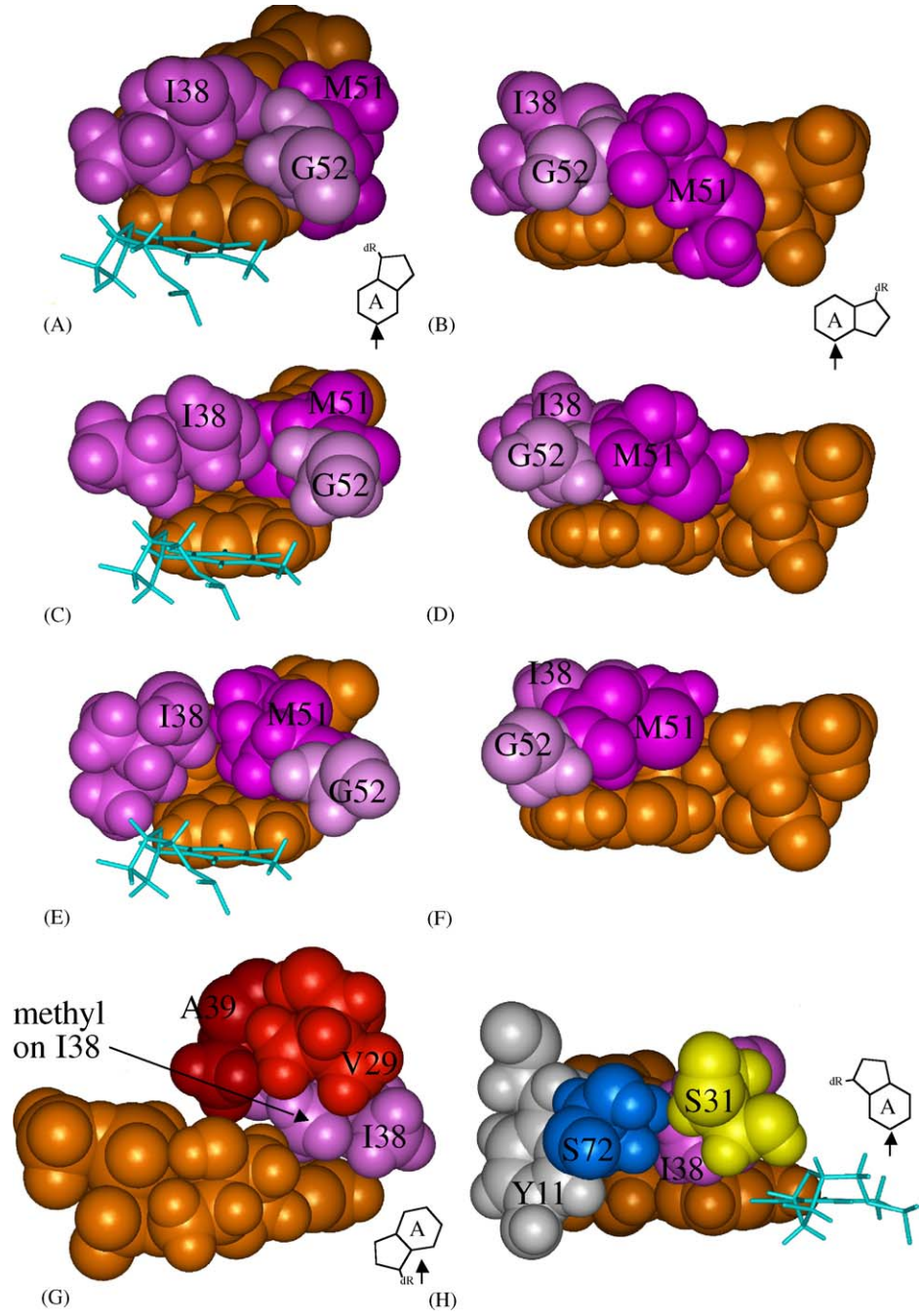

Fig. 4. UmuC amino acids in the vicinity of the adenine in dATP (brown) and the template thymine (turquoise). Arrows showing the relative viewing direction with respect to the adenine are included as inserts. (A) dATP (brown) with the three amino acids (purple) that touch dATP: roof-I38 (left, purple), G52 (center, light purple) and M51 (right/behind, dark purple). (B) View from the developing major groove of G52 (left, light purple) and M51 (right, dark purple). Panels (C) and (E) are analogous to panel (A), while panels (D) and (F) are analogous to panel (B), except with other UmuC structures (see text). (G) View of two amino acids V29 (right, red) and A39 (left, dark red) that form a pocket for the methyl group of the branched R-group of roof-I38 (purple). (H) View from the developing minor groove of S31 (yellow) and S72 (blue). Also included is Y11 (gray), which is close to I38 (purple) and stacks on the deoxyribose of dATP.

Fig. 5 shows a pictorial representation of this region from our models of UmuC, and DNAPs IV, η and κ, as well as from the X-ray structures of Dpo4 and DNAP ι. Each array in Fig. 5 includes the roof-aa at the center, surrounded by MaG1-aa, MaG2-aa, template-aa, MiG1-aa, MiG2-aa, dR1-aa and dR2-aa. Amino acid positioning aligns reasonable well, except for the amino acids in the developing minor groove (yellow, blue and white in Fig. 5). All of the DNAPs have an amino acid that aligns with V32 in Dpo4 (e.g. S31 = MiG1-aa (yellow) in UmuC, Fig. 4D). As shown in blue, S72 in UmuC aligns with M76 in Dpo4. In contrast, the amino acids in the position occupied by M76 of Dpo4 (and S72 of UmuC) are different in DNAP IV (G32) and DNAP κ (G131). Finally, in DNAP η, F18 occupies this position and it is from a completely different region of the protein.

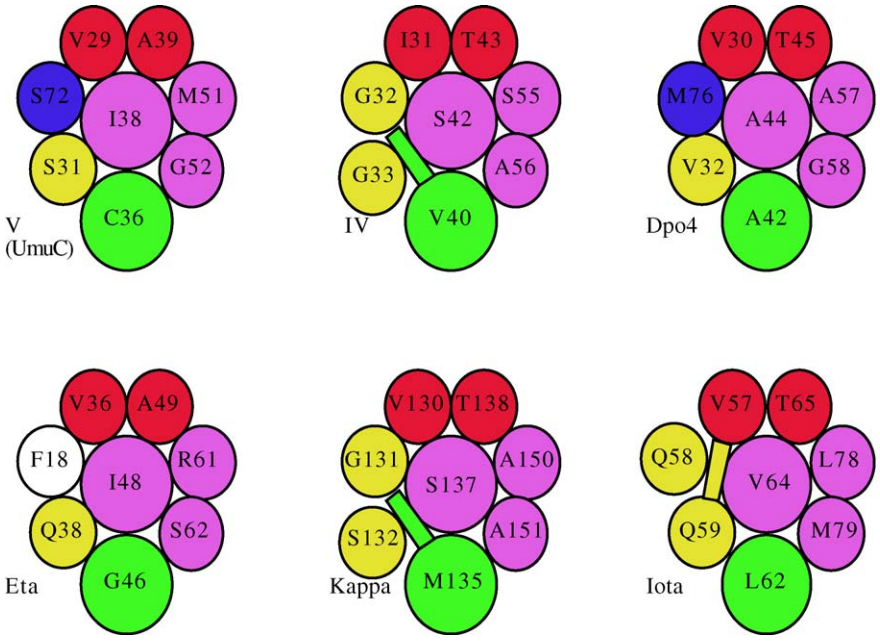

Fig. 5. Pictorial representation of the amino acids in the vicinity of the adenine in dATP in our models of UmuC, DNAP η, DNAP IV and DNAP κ, as well as from the X-ray coordinates of Dpo4 and DNAP ι. Color coding: purple for aa that contact adenine in dATP; green for the amino acid above template-dT; red for the amino acids that contact the roof-aa and lie above the deoxyribose of dATP; yellow, blue and white for amino acids that sit in the developing minor groove. The roof-aa is the larger/purple circle, while amino acids that sit in the developing major groove are smaller/purple circles. The rectangles indicate places where an amino acid blocks a second amino acid from gaining access to the roof-aa.

The greatest uncertainty in our models is undoubtedly the amino acids in the developing minor groove (MiG1 and MiG2). This uncertainty can be traced to the fact that this portion of each protein is formed from the confluence of three regions, as shown in Fig. 6 for UmuC. (This region looks very similar for Dpo4.) One region is made up of amino acids in a loop (∼aa27–38), which contains the roof-aa (I38), and the template-aa (C36), as well as one amino acid in the developing minor groove (S31). The second region comes from a β strand that includes approximately aa68–74 and provides S72 (blue in Fig. 6A). The third region is an α helix from approximately aa9–19, which includes a conserved aromatic amino acid (Y11, Fig. 6A) that face-stacks with the deoxyribose in the dNTP. In the case of our model of DNAP η, the equivalent amino acid (F18) not only retains its face-stacking with the deoxyribose, but it slides into contact with the roof-aa (Fig. 5). This seems to happen because Q38 in DNAPs η is bulky and displaces V80 (equivalent to S72 in UmuC), and thus allows room for F18 to

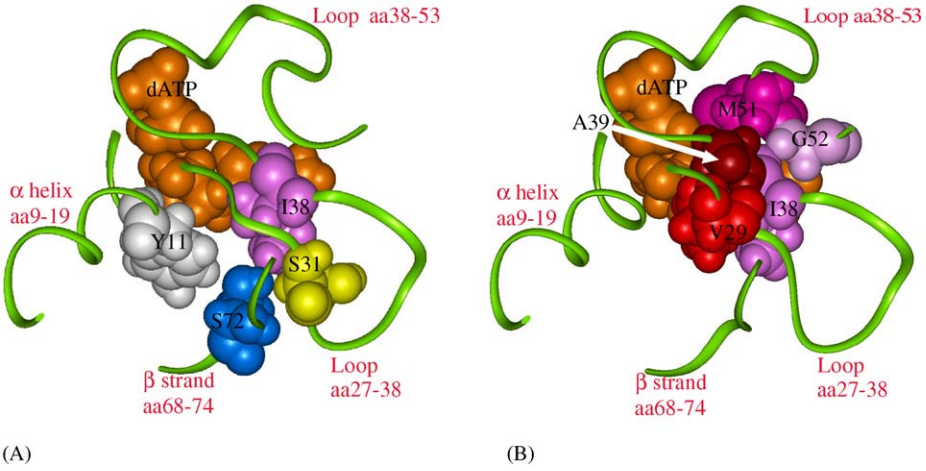

Fig. 6. Four regions (green ribbon) that contribute structure to UmuC in the vicinity of the adenine in dATP: loop (aa38–53), loop (aa27–38), α helix (aa9–19) and β strand (aa68–74). (A) Amino acids that bind in the developing minor groove (S31, yellow; S72, blue), along with the roof-isoleucine (I38), which sits on the adenine in dATP (brown), including its deoxyribose. Also included is Y11, which is close to the roof-I38 and stacks on the deoxyribose of dATP. (B) Amino acids that bind in the developing major groove (M51, dark purple; G52, light purple) and the roof-alanine (I38, purple), along with the amino acids that sit above the deoxyribose (V29, red; A39, dark red). The view is looking down on "roof-I38" as if perched above it.

slide in. This is easy to envision, given that in the equivalent position of UmuC, Y11 is close to I38 (Fig. 6A). (Similarly, Y12 is close to A44 in Dpo4.)

4.5. Comparing and contrasting the amino acids around the roof-amino acid in DNAPs IV, V, η and κ

Based on the pictorial representation in Fig. 5, the following observations can be made about similarities and differences of the amino acids in the vicinity of the dNTP binding pocket in DNAPs IV, V, η and κ.

(1) In terms of lesion-bypass, DNAP V (UmuC) and hDNAP η seem functionally related (Table 2). Concerning the amino acids that contact the base in the dNTP or the template, they have only one amino acid in common: I38 in UmuC, which is I48 in hDNAP η (Table 3, Fig. 5). I60 is the equivalent residue in yeast DNAP η, and in the X-ray structure it appears to occupy a similar position (Fig. 2A [33]), although DNA is not present for absolute confirmation. In contrast, other Y-family DNAPs (Dpo4, Dbh, IV, κ and ι) do not have isoleucine at the I38-position (Fig. 1). Furthermore, DNAPs IV and κ, which share dNTP insertional similarities (Table 2) both have serine in the roof position, and share one other amino acid in this region of their active site (MaG2-aa, Table 3). MaG1-aa has greater bulk in DNAPs V and η (M51 and R61, respectively) in comparison to in DNAPs IV and κ (S55 and A150, respectively). The side chain is non-bulky at MaG2-aa in all cases, except DNAP ι. Although DNAPs V and η have more bulky amino acids at the roof-aa and MaG1-aa positions, they have less bulk at the template-aa (C36 and G46, respectively) in comparison to DNAPs IV and κ (V40 and M135, respectively). In summary, the Y-family functional pairs seem to sort in terms of amino acids near the adenine of the dATP: (i) DNAPs V and η have a bulky isoleucine as the roof-aa and a bulky MaG1-aa, while the template-aa is non-bulky and (ii) DNAPs IV and κ have a non-bulky serine as the roof-aa and a non-bulky MaG1-aa, while the template-aa is bulky.

(2) Isoleucine has both a methyl group and an ethyl group in its branched R-group. In UmuC and DNAP η, the methyl group of the roof-isoleucine sits in a pocket where it contacts not only the adenine, but also the deoxyribose of dATP, as well as the conserved "red" amino acids (Figs. 4G, 5 and 6B). The methyl group on the roof-isoleucine in UmuC and DNAP η contacts a methyl group of A39 and A49, respectively, as well as an isopropyl group in V29 and V36, respectively (Fig. 4G). The less bulky roof-serines in DNAPs IV and κ interact with bulky "red" threonines (i.e. T43 and T138, respectively), in addition to having a bulky R-group at the other "red" amino acid (i.e. I31 and V130, respectively).

(3) Of our modeled structures derived from Dpo4, only S72 in UmuC retains the structural alignment with M76 in Dpo4 (blue in Fig. 5). A transformation occurred in the case of DNAP IV during the MD run, such that consecutive amino acids are now present as the minor groove representatives (i.e. G33/G32 for MiG1/MiG2 in DNAP IV). This transformation means that DNAP IV, which started off resembling Dpo4, was transformed during the MD run into a structure that resembles DNAP κ, which also has consecutive amino acids (i.e. S132/G131 for MiG1/MiG2). This similarity is apparent in a second way. The template-aa in DNAPs IV and κ is bulky (V40 and M135, respectively), which deforms the backbone slightly and prevents G33 in DNAP IV and S132 in DNAP κ from interacting with the roof-aa (Fig. 5).

Table 3
Amino acids in DNAPs V, η, IV and κ that interact with either the adenine in dATP or its complementary dT

Amino acid[a]	V	η	IV	κ
Roof-aa	I38	I48	S42	S137
MG1-aa	M51	R61	S55	A150
MG2-aa	G52	S62	A56	A151
Template-aa	C36	G46	V40	M135

[a] Amino acids that occupy the "roof-aa" position and sit on top of the adenine of dATP, or that occupy the "template-aa" position and sit on the complementary dT in the template, or that sit in the developing major groove ("MG1-aa" and "MG2-aa") and interact with both the adenine of dATP and the roof-aa.

In summary, there are a variety of structural similarities between the pairs UmuC/DNAP η and DNAPs IV/κ that could be important to their functional similarities (Table 2), although the true significance of any amino acid identity to dNTP insertion can only be assessed experimentally.

5. Summary

Homology modeling was used to construct starting coordinates for *E. coli* DNAPs IV and V, as well as human DNAPs η and κ, which were refined using molecular dynamics with CHARMM 27. A survey of the literature revealed that DNAPs V and η show similar patterns of dNTP insertion opposite a variety of DNA adducts/lesions (Table 2), suggesting they might be functional orthologs. DNAPs IV and κ also show a similar pattern with these same DNA lesions (Table 2), suggesting they too are functional orthologs. The active sites in our models of the four DNAPs were inspected for amino acids that might be relevant to the patterns of dNTP insertion opposite lesions (Fig. 5). Several patterns were noted, perhaps the most striking being that DNAPs V and η both have an isoleucine that stacks on top of the base of the incoming dNTP ("roof-aa"), while DNAPs IV and κ both have a serine in this same position. Although the relationship between these observations and dNTP insertional pattern and mechanism is not clear, these models suggest experiments to investigate the mechanism of dNTP choice/specificity during lesion-bypass experiments.

Acknowledgement

This work was supported by United States Public Health Services Grant R01CA50432 and R01ES03775.

References

[1] K. Bebenek, T.A. Kunkel, Functions of eukaryotic DNA polymerases, Adv. Prot. Chem. 69 (2004) 137–165.

[2] M.F. Goodman, Error-prone repair DNA polymerases in prokaryotes and eukaryotes, Annu. Rev. Biochem. 71 (2002) 17–50.

[3] U. Hubscher, G. Maga, S. Spadari, Eukaryotic DNA polymerases, Annu. Rev. Biochem. 71 (2002) 133–163.

[4] A.J. Rattray, J.N. Strathern, Error-prone DNA polymerases: when making a mistake is the only way to get ahead, Annu. Rev. Genet. 37 (2003) 31–66.

[5] H. Ohmori, E.C. Friedberg, R.P. Fuchs, M.F. Goodman, F. Hanaoka, D. Hinkle, T.A. Kunkel, C.W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G.C. Walker, Z. Wang, R. Woodgate, The Y-family of DNA polymerases, Mol. Cell. 8 (2001) 7–8.

[6] F. Boudsocq, H. Ling, W. Yang, R. Woodgate, Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass, DNA Repair (Amst.) 1 (2002) 343–358.

[7] E.C. Friedberg, R. Wagner, M. Radman, Specialized DNA polymerases, cellular survival, and the genesis of mutations, Science 296 (2002) 1627–1630.

[8] E.C. Friedberg, Why do cells have multiple error-prone DNA polymerases? Environ. Mol. Mutagen. 38 (2001) 105–110.

[9] Z. Wang, Translesion synthesis by the UmuC family of DNA polymerases, Mutat. Res. 486 (2001) 59–70.

[10] R.E. Johnson, M.T. Washington, S. Prakash, L. Prakash, Bridging the gap: a family of novel DNA polymerases that replicate faulty DNA, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 12224–12226.

[11] V. Pages, R.P. Fuchs, How DNA lesions are turned into mutations within cells? Oncogene 21 (2002) 8957–8966.

[12] W. Yang, Damage repair DNA polymerases Y, Curr. Opin. Struct. Biol. 13 (2003) 23–30.

[13] P.M. Burgers, E.V. Koonin, E. Bruford, L. Blanco, K.C. Burtis, M.F. Christman, W.C. Copeland, E.C. Friedberg, F. Hanaoka, D.C. Hinkle, C.W. Lawrence, M. Nakanishi, H. Ohmori, L. Prakash, S. Prakash, C.A. Reynaud, A. Sugino, T. Todo, Z. Wang, J.C. Weill, R. Woodgate, Eukaryotic DNA polymerases: proposal for a revised nomenclature, J. Biol. Chem. 276 (2001) 43487–43490.

[14] H. Ohmori, E.C. Friedberg, R.P. Fuchs, M.F. Goodman, F. Hanaoka, D. Hinkle, T.A. Kunkel, C.W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G.C. Walker, Z. Wang, R. Woodgate, The Y-family of DNA polymerases, Mol. Cell. 8 (2001) 7–8.

[15] C. Masutani, R. Kusumoto, A. Yamada, N. Dohmae, M. Yokoi, M. Yuasa, M. Araki, S. Iwai, K. Takio, F. Hanaoka, The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta, Nature 399 (1999) 700–704.

[16] R.E. Johnson, S. Prakash, L. Prakash, Efficient bypass of a thymine–thymine dimer by yeast DNA polymerase, Poleta, Science 283 (1999) 1001–1004.

[17] R.E. Johnson, C.M. Kondratick, S. Prakash, L. Prakash, hRAD30 mutations in the variant form of xeroderma pigmentosum, Science 285 (1999) 263–265.

[18] M.T. Washington, R.E. Johnson, S. Prakash, L. Prakash, Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta, J. Biol. Chem. 274 (1999) 36835–36838.

[19] M.T. Washington, R.E. Johnson, S. Prakash, L. Prakash, Accuracy of thymine–thymine dimer bypass by *Saccharomyces cerevisiae* DNA polymerase eta, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 3094–3099.

[20] R.E. Johnson, M.T. Washington, S. Prakash, L. Prakash, Fidelity of human DNA polymerase eta, J. Biol. Chem. 275 (2000) 7447–7450.

[21] O. Rechkoblit, Y. Zhang, D. Guo, Z. Wang, S. Amin, J. Krzeminsky, N. Louneva, N.E. Geacintov, Translesion synthesis past bulky benzo[*a*]pyrene diol epoxide N^2-dG and N^6-dA lesions catalyzed by DNA bypass polymerases, J. Biol. Chem. 277 (2002) 30488–30494.

[22] N. Suzuki, E. Ohashi, A. Kolbanovskiy, N.E. Geacintov, A.P. Grollman, H. Ohmori, S. Shibutani, Translesion synthesis by human DNA polymerase kappa on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-*anti*-N(2)-BPDE (7,8-dihydroxy-*anti*-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene), Biochemistry 41 (2002) 6100–6106.

[23] X. Huang, A. Kolbanovskiy, X. Wu, Y. Zhang, Z. Wang, P. Zhuang, S. Amin, N.E. Geacintov, Effects of base sequence context on translesion synthesis past a bulky (+)-*trans*-*anti*-B[*a*]P-N^2-dG lesion catalyzed by the Y-family polymerase pol kappa, Biochemistry 42 (2003) 2456–2466.

[24] Y. Zhang, X. Wu, D. Guo, O. Rechkoblit, Z. Wang, Activities of human DNA polymerase kappa in response to the major benzo[*a*]pyrene DNA adduct: error-free lesion bypass and extension synthesis from opposite the lesion, DNA Repair (Amst.) 1 (2002) 559–569.

[25] T. Ogi, Y. Shinkai, K. Tanaka, H. Ohmori, Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[*a*]pyrene, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 15548–15553.

[26] S. Avkin, M. Goldsmith, S. Velasco-Miguel, N. Geacintov, E.C. Friedberg, Z. Livneh, Quantitative analysis of translesion DNA synthesis across a benzo[*a*]pyrene-guanine adduct in mammalian cells: the role of DNA polymerase kappa, J. Biol. Chem. 279 (2004) 53298–53305.

[27] B.L. Zhou, J.D. Pata, T.A. Steitz, Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain, Mol. Cell. 8 (2001) 427–437.

[28] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication, Cell 107 (2001) 91–102.

[29] H. Ling, F. Boudsocq, B.S. Plosky, R. Woodgate, W. Yang, Replication of a *cis–syn* thymine dimer at atomic resolution, Nature 424 (2003) 1083–1087.

[30] H. Ling, J.M. Sayer, B.S. Plosky, H. Yagi, F. Boudsocq, R. Woodgate, D.M. Jerina, W. Yang, Crystal structure of a benzo[*a*]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 2265–2269.

[31] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts, Mol. Cell. 13 (2004) 751–762.

[32] J. Trincao, R.E. Johnson, W.T. Wolfle, C.R. Escalante, S. Prakash, L. Prakash, A.K. Aggarwal, Dpo4 is hindered in extending a G.T mismatch by a reverse wobble, Nat. Struct. Mol. Biol. 11 (2004) 457–462.

[33] J. Trincao, R.E. Johnson, C.R. Escalante, S. Prakash, L. Prakash, A.K. Aggarwal, Structure of the catalytic core of *S. cerevisiae* DNA polymerase eta: implications for translesion DNA synthesis, Mol. Cell. 8 (2001) 417–426.

[34] D.T. Nair, R.E. Johnson, S. Prakash, L. Prakash, A.K. Aggarwal, Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing, Nature 430 (2004) 377–380.

[35] S.N. Uljon, R.E. Johnson, T.A. Edwards, S. Prakash, L. Prakash, A.K. Aggarwal, Crystal structure of the catalytic core of human DNA polymerase kappa, Structure (Camb.) 12 (2004) 1395–1404.

[36] M.T. Washington, S.A. Helquist, E.T. Kool, L. Prakash, S. Prakash, Requirement of Watson–Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase eta, Mol. Cell. Biol. 23 (2003) 5107–5112.

[37] M.T. Washington, L. Prakash, S. Prakash, Mechanism of nucleotide incorporation opposite a thymine–thymine dimer by yeast DNA polymerase eta, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 12093–12098.

[38] J. Yin, K.-Y. Seo, E.L. Loechler, A role for DNA polymerase V in G → T mutagenesis from the major benzo[*a*]pyrene N^2-dG adduct when studied in a 5′-TGT sequence in *Escherichia coli*, DNA Repair 3 (2004) 323–334.

[39] R.G. Harvey, Polycyclic Aromatic Hydrocarbons: Chemistry and Cancer, Wiley-VCH Inc., New York, 1997.

[40] D.H. Phillips, Fifty years of benzo[*a*]pyrene, Nature 303 (1983) 468–472.

[41] B. Singer, D. Grunberger, Molecular Biology of Mutagens and Carcinogens, Plenum Press, New York, 1983.

[42] A.H. Conney, Induction of microsomal enzymes by foreign chemicals and carcinogens by polycyclic aromatic hydrocarbons, Cancer Res. 42 (1982) 4875–4917.

[43] A. Dipple, Polycyclic aromatic hydrocarbon carcinogens, in: R.G. Harvey (Ed.), Polycyclic Aromatic Hydrocarbons and Carcinogenesis, American Chemical Society Press, Washington, DC, 1985 , pp. 1–17.
[44] P.W. Jones, Polynuclear aromatic hydrocarbons, in: M.G. Bowman (Ed.), Handbook of Carcinogens and Hazardous Substances, Marcel Dekker Inc., New York, 1982, pp. 573–639.
[45] P. Grasso, Carcinogens in food, in: C.E. Searle (Ed.), Chemical Carcinogens, second ed., ACS Monograph 182, Amreican Chemical Society Press, Washington DC, 1984, pp. 1205–1239.
[46] J.M. Samet, D.M. DeMarini, H.V. Malling, Do airborne particles induce heritable mutations? Science 304 (2004) 971–972.
[47] A. Balmain, K. Brown, Oncogene activation in chemical carcinogenesis, Adv. Cancer Res. 51 (1988) 147–182.
[48] M. Hall, P.L. Grover, in: G.S. Cooper, P.L. Grover (Eds.), Handbook of Experimental Pharmacology, vol. 94/1 Chemical Carcinogenesis and Mutagenesis, Springer-Verlag, Heidelberg, 1990, pp. 327–372.
[49] B.A. Ruggeri, B. Bauer, S.-Y. Zhang, A.J.P. Klein-Szanto, Murine squamous cell carcinoma cell lines produced by a complete carcinogenesis protocol with benzo[*a*]pyrene exhibit characteristic p53 mutations and the absence of H-ras and cyl1/cyclin D1 abnormalities, Carcinogenesis 15 (1994), 6113-1619.
[50] D. Chakravarti, J. Pelling, E.L. Cavalieri, E.G. Rogan, Relating aromatic hydrocarbon-induced DNA adducts and the c-H-ras mutations in mouse skin papillomas: the role of apurinic sites, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 10422–10426.
[51] L. Chen, R. Devanesan, S. Higginbotham, F. Ariese, R. Jankowiak, G.J. Small, E.G. Rogan, E. Cavalieri, Expanded analysis of benzo[*a*]pyrene-DNA adducts formed in vitro and in mouse skin: their significance in tumor initiation, Chem. Res. Toxicol. 9 (1996) 897–903.
[52] D. Chakravarti, P. Mailander, J. Franzen, S. Higginbotham, E.L. Cavalieri, E.G. Rogan, Detection of dibenzo[*a*,l]pyrene-induced H-ras codon 61 mutant genes in preneoplastic SENCAR mouse skin using a new PCR-RFLP method, Oncogene 16 (1998) 3203–3210.
[53] D. Chakravarti, P.C. Mailander, E.L. Cavalieri, E.G. Rogan, Evidence that error-prone DNA repair converts dibenzo[*a*,l]pyrene-induced depurinating lesions into mutations: formation, clonal proliferation and regression of initiated cells carrying H-ras oncogene mutations in early preneoplasia, Mutat. Res. 456 (2000) 17–32.
[54] G.P. Pfeifer, P. Hainaut, On the origin of G → T transversions in lung cancer, Mutat. Res. 526 (2003) 39–43.
[55] E. Eisenstadt, A.J. Warren, J. Porter, D. Atkins, J.H. Miller, Carcinogenic epoxides of benzo[*a*]pyrene and cyclopenta[cd]pyrene induce base substitutions via specific transversions, Proc. Natl. Acad. Sci. U.S.A. 82 (1982) 1945–1949.
[56] J.L. Yang, V.M. Maher, J.J. McCormick, Kinds of mutations formed when a human shuttle vector containing adducts of (+/−)-7beta,8alpha-dihydroxy-9alpha,10alpha-epoxy-7,8,9,10- dihydrobenzo[*a*]pyrene replicates in human cells, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 3787–3791.
[57] C. Bernelot-Moens, B.W. Glickman, A.J.E. Gordon, Induction of specific frameshift and base substitution events by benzo[*a*]pyrene diol epoxide in excision-repair-deficient *Escherichia coli*, Carcinogenesis 11 (1990) 781–785.
[58] J.-L. Yang, R.-H. Chen, V.M. Maher, J.J. McCormick, Kinds and locations of mutations induced by (+/−)-7beta, 8alpha-dihydroxy-9alpha, 10alpha-epoxy-7,8,9,10-dihydrobenzo[*a*]pyrene in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in diploid human fibroblasts, Carcinogenesis 12 (1990) 71–75.
[59] A.M. Carothers, D. Grunberger, DNA base changes in benzo[*a*]pyrene diol epoxide-induced dihydrofolate reductase mutants of Chinese hampster ovary cells, Carcinogenesis 11 (1990) 189–192.
[60] H. Rodriguez, E.L. Loechler, Mutational spectra of the (+)-*anti*-diol epoxide of senzo[*a*]pyrene in a supF gene of an *Escherichia coli* plasmid: DNA sequence context influences hotspots mutational specficity and the extent of SOS enhancement of mutagenesis, Carcinogenesis 14 (1993) 373–383.
[61] H. Rodriguez, E.L. Loechler, Mutagenesis by the (+)-*anti*-diol epoxide of benzo[*a*]pyrene: what controls mutagenic specificity? Biochemistry 32 (1993) 373–383.
[62] S.J. Wei, R.L. Chang, C.Q. Wong, X.X. Cui, N. Dandamudi, Y.P. Lu, K.A. Merkler, J.M. Sayer, A.H. Conney, D.M. Jerina, The ratio of deoxyadenosine to deoxyguanosine adducts formed by (+)-(7*R*,8*S*,9*S*,10*R*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene in purified calf thymus DNA and DNA in V-79 cells is independent of dose, Int. J. Oncol. 14 (1999) 509–513.
[63] M. Schiltz, X.X. Cui, Y.P. Lu, H. Yagi, D.M. Jerina, M.Z. Zdzienicka, R.L. Chang, A.H. Conney, S.J. Wei, Characterization of the mutational profile of (+)-7*R*,8*S*-dihydroxy-9*S*, 10*R*-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene at the hypoxanthine (guanine) phosphoribosyltransferase gene in repair-deficient Chinese hamster V-H1 cells, Carcinogenesis 20 (1999) 2279–2286.
[64] W. Mackay, M. Benasutti, E. Drouin, E.L. Loechler, Mutagenesis by the major adduct of activated benzo[*a*]pyrene, (+)-*anti*-BP-N^2-Gua when studied in an *Escherichia coli* plasmid using site-directed methods, Carcinogenesis 13 (1992) 1415–1425.
[65] R. Shukla, N. Geacintov, E.L. Loechler, The major N^2-dG adduct of (+)-*anti*-B[*a*]PDE induces G → A mutations in a 5′-AGA-3′ sequence context, Carcinogenesis 20 (1999) 261–268.
[66] N. Lenne-Samuel, R. Janel-Bintz, A. Kolbanovskiy, N.E. Geacintov, R.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*, Mol. Microbiol. 38 (2000) 299–307.
[67] R. Napolitano, R. Janel-Bintz, J. Wagner, R.P. Fuchs, All three SOS-inducible DNA polymerases (Pol II Pol IV and Pol V) are involved in induced mutagenesis, EMBO J. 19 (2000) 6259–6265.
[68] R.L. Napolitano, I.B. Lambert, R.P. Fuchs, SOS factors involved in translesion synthesis, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 5733–5738.
[69] O.J. Becherel, R.P. Fuchs, Mechanism of DNA polymerase II-mediated frameshift mutagenesis, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 8566–8571.
[70] J. Wagner, H. Etienne, R. Janel-Bintz, R.P. Fuchs, Genetics of mutagenesis in *E. coli*: various combinations of translesion polymerases (Pol II IV and V) deal with lesion/sequence context diversity, DNA Repair 1 (2002) 159–167.
[71] D. Chiapperino, H. Kroth, I.H. Kramarczuk, J.M. Sayer, C. Masutani, F. Hanaoka, D.M. Jerina, A.M. Cheh, Preferential misincorporation of purine nucleotides by human DNA polymerase eta opposite benzo[*a*]-pyrene 7,8-diol, 9, 10-epoxide deoxyguanosine adducts, J. Biol. Chem. 277 (2002) 11765–11771.
[72] Y. Zhang, X. Wu, D. Guo, O. Rechkoblit, N.E. Geacintov, Z. Wang, Two-step error-prone bypass of the (+)- and (−)-*trans-anti*-BPDE-N(2)-dG adducts by human DNA polymerases eta and kappa, Mutat. Res. 510 (2002) 23–35.
[73] Y. Zhang, F. Yuan, X. Wu, O. Rechkoblit, J.S. Taylor, N.E. Geacintov, Z. Wang, Error-prone lesion bypass by human DNA polymerase eta, Nucleic Acids Res. 28 (2000) 4717–4724.
[74] X. Shen, J.M. Sayer, H. Kroth, I. Ponten, M. O'Donnell, R. Woodgate, D.M. Jerina, M.F. Goodman, Efficiency and accuracy of SOS-induced DNA polymerases replicating benzo[*a*]pyrene-7,8-diol 9,10-epoxide A and G adducts, J. Biol. Chem. 277 (2002) 5265–5274.
[75] Insight II, Version 98.0, Accelyrs Inc., San Diego, 1998.
[76] L. Nilsson, M. Karplus, Empirical energy functions for energy minimization and dynamics of nucleic acids, J. Comp. Chem. 7 (1986) 591–616.
[77] A.D. MacKerell Jr., N. Banavali, N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, Biopolymers 56 (2001) 257–265.
[78] S. Chandani, C.H. Lee, E.L. Loechler, Molecular modeling in water of the major benzo[*a*]pyrene N^2-dG adduct using free energy perturbation techniques to assess energy differences for two conformations in two sequence contexts, Chem. Res. Toxicol. 18 (2005) 1108–1123.

[79] S. Doublie, S. Tabor, A.M. Long, C.C. Richardson, T. Ellenberger, Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution, Nature 391 (1998) 251–258.
[80] J.U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164–170.
[81] D. Eisenberg, M. Wesson, M. Yamashita, Interpretation of protein folding and binding with atomic solvation parameters, Chem. Scr. 29A (1989) 217–221.
[82] M. Gribskov, A.D. McLachlan, D. Eisenberg, Profile analysis: detection of distantly related proteins, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 4355–4358.
[83] M. Gribskov, R. Lüthy, D. Eisenberg, Profile analysis, Methods Enzymol. 183 (1990) 146–159.
[84] S. Fujii, V. Gasser, R.P. Fuchs, The biochemical requirements of DNA polymerase V-mediated translesion synthesis revisited, J. Mol. Biol. 341 (2004) 405–417.
[85] S. Fujii, R.P. Fuchs, Defining the position of the switches between replicative and bypass DNA polymerases, EMBO J. 23 (2004) 4342–4352.
[86] M. Tang, P. Pham, X. Shen, J.S. Taylor, M. O'Donnell, R. Woodgate, M.F. Goodman, Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis, Nature 404 (2000) 1014–1018.
[87] J.H. Lee, G.S. Hwang, B.S. Choi, Solution structure of a DNA decamer duplex containing the stable 3′ T.G base pair of the pyrimidine(6-4)pyrimidone photoproduct [(6-4) adduct]: implications for the highly specific 3′ T → C transition of the (6-4) adduct, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6632–6636.
[88] R. Napolitano, R. Janel-Bintz, J. Wagner, R.P. Fuchs, All three SOS-inducible DNA polymerases (Pol II Pol IV and Pol V) are involved in induced mutagenesis, EMBO J. 19 (2000) 6259–6265.
[89] J. Wagner, H. Etienne, R. Janel-Bintz, R.P. Fuchs, Genetics of mutagenesis in *E. coli*: various combinations of translesion polymerases (Pol II, IV and V) deal with lesion/sequence context diversity, DNA Repair 1 (2002) 159–167.
[90] R.L. Napolitano, I.B. Lambert, R.P. Fuchs, SOS factors involved in translesion synthesis, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 5733–5738.
[91] O.J. Becherel, R.P. Fuchs, Mechanism of DNA polymerase II-mediated frameshift mutagenesis, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 8566–8571.
[92] B.S. Strauss, The "A" rule revisited: polymerases as determinants of mutational specificity, DNA Repair 1 (2002) 125–135.
[93] E.L. Loechler, Mechanism by Which Aflatoxins and Other Bulky Carcinogens Induce Mutations. The Toxicology of Aflatoxins: Human Health, Veterinary and Agricultural Significance (Eaton and Groopman), Academic Press, Orlando, 1994 , pp. 149–178 (Chapter 8).
[94] D.E. Brash, W.A. Haseltine, UV-induced mutation hotspots occur at DNA damage hotspots, Nature 298 (1982) 189–192.
[95] C. Masutani, R. Kusumoto, S. Iwai, F. Hanaoka, Mechanisms of accurate translesion synthesis by human DNA polymerase eta, EMBO J. 19 (2000) 3100–3109.
[96] A.M. Cordonnier, A.R. Lehmann, R.P. Fuchs, Impaired translesion synthesis in xeroderma pigmentosum variant extracts, Mol. Cell. Biol. 19 (1999) 2206–2211.
[97] Y. Zhang, F. Yuan, X. Wu, O. Rechkoblit, J.S. Taylor, N.E. Geacintov, Z. Wang, Error-prone lesion bypass by human DNA polymerase eta, Nucleic Acids Res. 28 (2000) 4717–4724.
[98] F. Yuan, Y. Zhang, D.K. Rajpal, X. Wu, D. Guo, M. Wang, J.S. Taylor, Z. Wang, Specificity of DNA lesion bypass by the yeast DNA polymerase eta, J. Biol. Chem. 275 (2000) 8233–8239.
[99] S.L. Yu, R.E. Johnson, S. Prakash, L. Prakash, Requirement of DNA polymerase eta for error-free bypass of UV-induced CC and TC photoproducts, Mol. Cell. Biol. 21 (2001) 185–188.
[100] N. Suzuki, E. Ohashi, K. Hayashi, H. Ohmori, A.P. Grollman, S. Shibutani, Translesional synthesis past acetylaminofluorene-derived DNA adducts catalyzed by human DNA polymerase kappa and *Escherichia coli* DNA polymerase IV, Biochemistry 40 (2001) 15176–15183.
[101] E. Ohashi, T. Ogi, R. Kusumoto, S. Iwai, C. Masutani, F. Hanaoka, H. Ohmori, Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa, Genes Dev. 14 (2000) 1589–1594.
[102] Y. Zhang, F. Yuan, X. Wu, M. Wang, O. Rechkoblit, J.S. Taylor, N.E. Geacintov, Z. Wang, Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro, Nucleic Acids 28 (2000) 4138–4146.
[103] V.L. Gerlach, W.J. Feaver, P.L. Fischhaber, E.C. Friedberg, Purification and characterization of pol kappa, a DNA polymerase encoded by the human DINB1 gene, J. Biol. Chem. 276 (2001) 92–98.
[104] R.E. Johnson, S. Prakash, L. Prakash, The human DINB1 gene encodes the DNA polymerase Poltheta, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 3838–3843.
[105] E. Ohashi, K. Bebenek, T. Matsuda, W.J. Feaver, V.L. Gerlach, E.C. Friedberg, H. Ohmori, T.A. Kunkel, Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene, J. Biol. Chem. 275 (2000) 39678–39684.
[106] E. Ohashi, T. Ogi, R. Kusumoto, S. Iwai, C. Masutani, F. Hanaoka, H. Ohmori, Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa, Genes Dev. 14 (2000) 1589–1594.
[107] Y. Zhang, F. Yuan, X. Wu, O. Rechkoblit, J.S. Taylor, N.E. Geacintov, Z. Wang, Error-prone lesion bypass by human DNA polymerase eta, Nucleic Acids Res. 28 (2000) 4717–4724.
[108] P. Prevelige Jr., G. Fasman, in: G. Fasman (Ed.), Prediction of Protein Structure and the Principles of Protein Conformation, Plenum, New York, 1989 , pp. 391–416 (Chapter 9).
[109] J. Garnier, B. Robson, in: G. Fasman (Ed.), Prediction of Protein Structure and the Principles of Protein Conformation, Plenum, New York, 1989 , pp. 417–465 (Chapter 10).
[110] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.