ELSEVIER

DNA Repair 1 (2002) 343–358

DNA
REPAIR

www.elsevier.com/locate/dnarepair

Mini review

Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass

François Boudsocq [a], Hong Ling [b], Wei Yang [b], Roger Woodgate [a,*]

[a] Section on DNA Replication, Repair and Mutagenesis, Building 6, Room 1A13, National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892-2725, USA

[b] Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0510, USA

Accepted 7 March 2002

Abstract

Our understanding of the molecular mechanisms of error-prone lesion bypass has changed dramatically in the past few years. The concept that the key participants in the mutagenic process were accessory proteins that somehow modified the ability of the cell's main replicase to facilitate bypass of normally blocking lesions has been replaced with one in which the replicase is displaced by a polymerase specialized in lesion bypass. The participants in this process remain the same, only their function has been reassigned. What was once known as the UmuC/DinB/Rev1/Rad30 superfamily of mutagenesis proteins, is now known as the Y-family of DNA polymerases. Quite remarkably, within the space of 3 years, the field has advanced from the initial discovery of intrinsic polymerase function, to the determination of the tertiary structures of several Y-family DNA polymerases.

A key to determining the biochemical properties of each DNA polymerase is through structure–function studies that result in the site-specific substitution of particular amino acids at critical sites within each DNA polymerase. However, we should not forget the power of genetic selection that allows us to identify residues within each polymerase that are generated by "random mutagenesis" and which are important for both a gain or loss of function in vivo. In this review, we discuss the structural ramifications of several missense mutations previously identified in various Y-family DNA polymerase and speculate on how each amino acid substitution might modify the enzymatic activity of the respective polymerase or possibly perturb protein–protein interactions necessary for efficient translesion replication in vivo. Published by Elsevier Science B.V.

Keywords: DNA replication; pol IV; pol V; pol η; pol ι; pol κ; Rev1

1. Introduction: an historical perspective to the Y-family DNA polymerases

Ever since the isolation of non-mutable *REV* strains of *Saccharomyces cerevisiae* [1] or *umu* strains of *Escherichia coli* [2,3] some 25–30 years ago, we have known that damage-induced mutagenesis is not a passive process, but rather one that actively requires the participation of several key proteins. A key link between prokaryotes and eukaryotes in the mutagenic process was forged when Larimer et al., cloned and sequenced the *S. cerevisiae REV1* gene and reported that the encoded Rev1 protein exhibited 25% identity and 42% similarity to *E. coli* UmuC over a stretch

* Corresponding author. Tel.: +1-301-496-6175;
fax: +1-301-594-1135.
E-mail address: woodgate@helix.nih.gov (R. Woodgate).

1568-7864/02/$ – see front matter. Published by Elsevier Science B.V.
PII: S1568-7864(02)00019-8

of 152 amino acids located within its N-terminus [4]. Around the same time, several orthologs were cloned and sequenced from various enterobacteriaceae [5–9] and their naturally occurring R-plasmids [10–13], as well as from the archaeon *Sulfolobus solfataricus* P1 [14], suggesting the existence of a so-called family of "mutagenesis proteins" [14]. With the advent of entire genome analysis, the number of orthologs has increased dramatically. Database searches now reveal well over 100 orthologs in all three kingdoms of life. Phylogenetic analysis of the proteins gives rise to an unrooted tree consisting of several discrete sub-branches that are best typified by the *E. coli* UmuC and DinB proteins and the *S. cerevisiae* Rev1 and Rad30 proteins [15–17].

Based upon genetic and physiological experiments, it was generally believed that members of the superfamily were replication accessory factors which somehow relaxed the high fidelity of the cells replicase and in doing so, coerced the polymerase to extend nascent DNA primers that had been errantly incorporated opposite the normally replication-blocking lesion [18,19].

The paradigm began to evolve however, in 1996 when Nelson et al. [20], demonstrated that *S. cerevisiae* Rev1 protein exhibited deoxycytidyl transferase activity and is able to incorporate dCMP directly opposite an abasic lesion, thereby implying that the "mutagenesis proteins" might actually play a more active role than previously thought. Indeed, support for such a notion came a few years later when Johnson et al., overproduced and purified the *S. cerevisiae* Rad30 protein. Rad30 had previously been identified and characterized as belonging to the "mutagenesis" family based upon its similarity to *E. coli* UmuC and DinB as well as *S. cerevisiae* Rev1 proteins [21,22], but quite surprisingly, characterization of the purified protein indicated that unlike Rev1, Rad30 actually utilizes all four nucleoside triphosphates in a template-dependent manner and, as a consequence, meets all of the criteria of a *bona fide* DNA polymerase [23]. Remarkably, within a period of a few short months, similar studies demonstrated that both the *E. coli dinB* and *umuC* genes also encode DNA polymerases, termed pols IV and V, respectively [24–26], as do the human *RAD30A*, *RAD30B* and *DINB1* genes which encode human pols η, ι and κ, respectively [27–35]. The metamorphosis from a family of "mutagenesis proteins" to a family of DNA polymerases recently concluded when it was suggested that proteins phylogenetically related to UmuC/DinB/Rev1/Rad30 be called the "Y-Family" of DNA polymerases [17].

The fact that the proteins possess intrinsic DNA polymerase activity is quite surprising as there is little, to no, similarity in the primary amino acid sequence of Y-family polymerases with those previously identified from the A-, B-, C-, D-, or X-polymerase families [36–38]. In the past few months, however, the tertiary structures of several Y-family polymerases have been solved by X-ray crystallography [39–42] and their structures found to be remarkably similar to those previously reported for other DNA polymerases [43–47]. Like polymerases from the A-, B-, C- and X-families [48], the Y-family polymerases posses a topology akin to a right hand, with "thumb", "fingers" and "palm" domains [39–41]. The Y-family polymerases differ however, from high fidelity polymerases in that they contain an additional domain that has been termed the "little finger" [41]; the polymerase-associated domain (PAD) [40]; or the "wrist" [42].

Enzymatic characterization of the Y-family DNA polymerases in vitro reveals that on undamaged DNA templates, they are distributive and that DNA synthesis is inherently error-prone [24–26,29–32,35,42,49–53]. In contrast, their ability to facilitate lesion bypass appears to be polymerase and often sequence context specific [29,33,54–58]. Interestingly, recent studies suggest that both the processivity of the Y-family polymerases on undamaged DNA templates and their ability to bypass replication-blocking lesions is strongly influenced by certain protein cofactors. Perhaps the best example is that of the *E. coli* UmuC protein, the polymerase activity of which is dramatically increased in the presence of UmuD′ [26], with which it is normally complexed [59,60]) and through interactions with the RecA protein [25,54,61,62]. Just as the activity of *E. coli* pol V is stimulated via interactions with UmuD′ and RecA, it seems reasonable to expect that the enzymatic activity of other Y-family polymerases will be modulated though interactions with yet to be identified protein-partners. Recent studies suggest that one of these protein-partners is likely to include the β or proliferating cell nuclear antigen (PCNA)-like sliding clamp which together

with its appropriate clamp loader (*E. coli* γ-complex or eukaryotic RFC complex), has been shown to stimulate the catalytic activity of *E. coli* pols IV and V [25,54,61–63], *S. solfataricus* P1 Dbh (pol Y1) [53], as well as *S. cerevisiae* pol η and human pols η, ι and κ [64–67], indicating that such interactions may be universal within the Y-family polymerases.

The structural determination of several Y-family polymerases has resulted in significant advances in understanding the molecular basis for their low-fidelity synthesis on undamaged DNA and concomitant ability to bypass DNA lesions. In the ∼30 years intervening their initial genetic identification and their subsequent structural determination, the Y-family polymerases have been the subject of a variety of genetic analyzes which resulted in the isolation of a number of mutants encoding proteins with either a gain or loss of function. Here, we attempt to explain the basis for these phenotypes by mapping the location of several missense mutations previously identified in *E. coli* UmuC and DinB, and *S. cerevisiae* and human pol η, onto the known structure of the *S. solfataricus* Dpo4 protein [41] or the catalytic core of *S. cerevisiae* pol η [40]. In some cases, it appears that the mutations are located in key structural residues, while others effect the ability of the polymerase to bind DNA or an incoming nucleotide. However, a few mutants are located on the surface of the protein and we hypothesize that these residues may possibly represent regions of contact between Y-family polymerases and their protein-partners.

2. Discussion

The four manuscripts describing the structure of *S. solfataricus* P1 Dbh protein [39,42], *S. solfataricus* P2 Dpo4 protein [41] and the *S. cerevisiae* Rad30 protein [40] each contained an alignment of several Y-family polymerases that was generated based upon their primary amino acids sequences. Now that the tertiary structures of three Y-family polymerases have been determined, we have refined the alignment based upon the conserved structure of the polymerases, rather than their primary amino acid sequence and this new alignment is shown in Fig. 1. This figure provides the basis for determining which residue is altered in the orthologous polymerase.

2.1. Mutations in E. coli UmuC that reduce SOS mutagenesis

The first mutations in *umuC* were identified in the late 1970s by Kato and Shinoura (*umuC36*) [2] and independently by Steinborn (*umuC25* and *umuC104*) [3] in genetic screens for *E. coli* that were non-mutable after exposure to ultraviolet light. All three alleles were subsequently shown to result from single base-pair substitutions that generate missense mutations [68]. Perhaps the most interesting is *umuC104* (UmuC; D101N). The fact that the UmuC104 protein is relatively stable in vivo, suggested that its inability to promote damage-induced mutagenesis might lie in a catalytic rather than structural defect in the protein [69]. Indeed, we now know that UmuC D101 is one of three carboxylate residues (UmuC D6 and E102 being the others) that are absolutely conserved in the Y-family polymerases and constitute the active site of the enzyme (Fig. 2). In particular D6 and D101, like their counterparts in Dpo4 (D7 and D105, respectively) are predicted to chelate the divalent metal ion required for catalysis [41]. Substitution of Asp101 for Asn will disrupt binding of the cation, which explains why the mutant UmuC104 protein is unable to facilitate translesion replication in vitro [25,26].

UmuC36 has an E75K substitution in UmuC [68]. The mutation generally renders *E. coli* non-mutable [2,69]. However, Bates et al. [70], demonstrated that overproduction of UmuD′ could partially restore UV-induced mutagenesis in excision-defective strains of *E. coli* [70], suggesting that the mutant UmuC36 protein may have a reduced ability to interact with its cognate partner, UmuD′ [70]. E75 in *E. coli* UmuC corresponds to E79 in Dpo4, which is located at the N-terminus of helix D, near the junction between the finger and palm domains of the polymerase. Of especial interest is the observation that E79 is on the surface of the polymerase and is exposed to the solvent groove which separates the conserved fingers, palm and thumb domains from the little finger domain (Fig. 3). The location of this residue and the phenotype of the mutant protein suggests, therefore, that E75 and its surrounding/residues may provide an interface for UmuC to potentially interact with UmuD′.

The *umuC25* allele results in a T290K substitution within UmuC [68]. We believe that UmuC T290 corresponds to A288 in Dpo4. This residue is located in

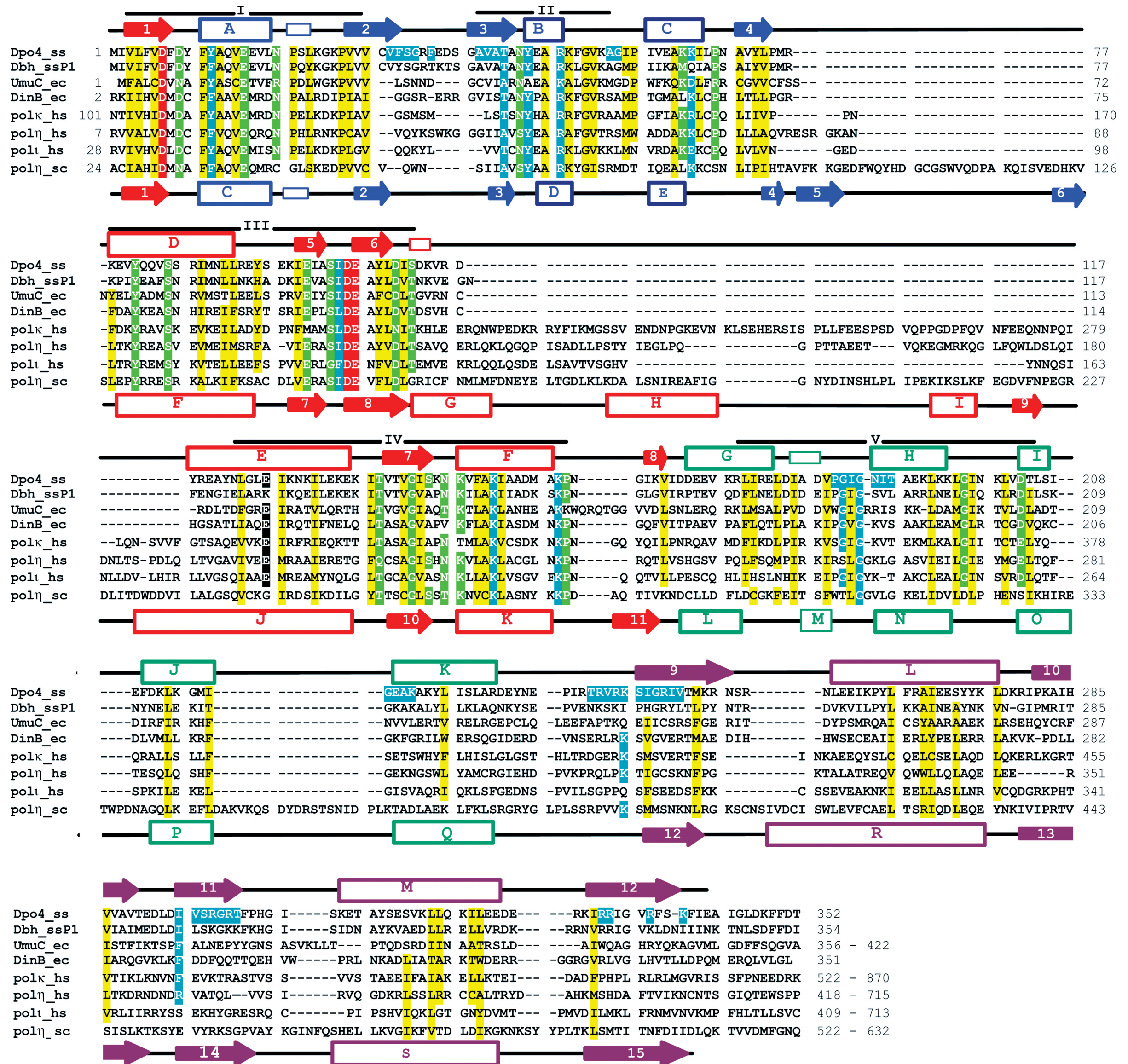

Fig. 1. Revised structure-based sequence alignment of Y-family polymerases. *S. solfataricus* strain P2 Dpo4 (Dpo4_ss, AAK42588), *S. solfataricus* strain P1 Dbh (Dbh_ssP1, T46875), *E. coli* UmuC/pol V (UmuC_ec, AAA24729), *E. coli* DinB/pol IV (DinB_ec, Q47155), *Homo sapiens* pol κ/DINB1 (pol κ_hs, XP_048874), *H. sapiens* pol η/Rad30A (pol η_hs, AF158185_1), *H. sapiens* pol ι/Rad30B (pol ι_hs, AF140501_1) and *S. cerevisiae* pol η/Rad30, (pol η_sc, 18158626) are included. The revised alignment is based on the published structures of *S. solfataricus* P1 Dbh, *S. solfataricus* P2 Dpo4 and *S. cerevisiae* pol η [39–42]. Secondary structures are indicated above the aligned sequence for Dpo4 and Dbh and under the alignment for *S. cerevisiae* pol η. The four structural domains are colored as in Ling et al. [41], where red is the palm domain; blue the fingers domain; green, the thumb domain and purple, the little finger domain. The α-helices are shown as rectangular boxes, β-strands as arrows and 3_{10} helices as narrow rectangles. The black line terminates at the end of the known structure of the polymerases. In Dpo4 this is at residue 341, in Dbh at residue 345 and in pol η at residue 509. Conserved residues important for the hydrophobic core formation are highlighted in yellow; those important for structural integrity in green; and those for polymerase activity in red. Residues interacting with DNA are highlighted in blue, whether conserved or not. E127 highlighted in black appears to be conserved, but it is completely exposed to solvent. The previously identified five sequence motifs [72] are indicated above the alignment.

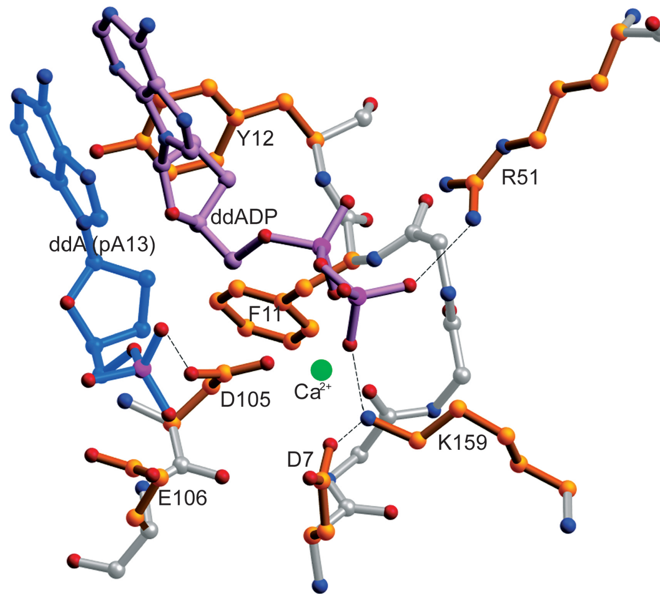

Fig. 2. Close up of the active site of Dpo4. Residues important for Dpo4's catalytic activity are shown along with the Ca^{2+} ion (green), the 3′ end nucleotide of the primer strand pA13 (ddA, blue) and the incoming nucleotide (ddADP, pink). The carbon backbone of residues in Dpo4 that frame the active site is shown in gray and their side chains in orange. Hydrogen bonds involved in holding the active site together are drawn in dashed lines. This figure was generated using ribbons [101].

the hydrophobic core of the little finger domain and introduction of a highly charged side chain, such as Lys, is predicted to disturb the stability of the little finger domain. In agreement with this hypothesis, the mutant UmuC25 protein is extremely unstable in vivo and is rapidly degraded [69].

In 1994, Woodgate et al., utilized a papillation assay to identify a number of *umuC* mutants that had a reduced ability to promote spontaneous SOS mutagenesis [69]. The phenotype of many of these mutants could be attributed to a change in the expression of UmuC (either an increase or decrease), but several alleles were also expressed at close to normal cellular levels, suggesting that they may harbor defects in catalytic activity or protein–protein interactions necessary for efficient SOS mutagenesis. One of these was a mutant in which R279 was substituted with cysteine [69]. The mutant protein exhibited no SOS-dependent spontaneous mutagenesis and a very low level of methyl methane sulphonate (MMS)-induced mutagenesis. This residue corresponds to D278 in Dpo4 and is located in a solvent exposed loop of the "little finger" domain and is on the back side of the polymerase "hand" away from DNA binding sites (Fig. 3). It is unlikely that the R → C substitution causes any gross structural changes in the polymerase. UmuC R279 is predicted to be in close proximity to E75, but on the opposite side of the solvent exposed groove (Fig. 3). It is therefore possible that this entire region represents an interface through which UmuC might interact with UmuD′. Against this notion is the observation that the R279C mutant protein exhibits wild-type stability [69]. UmuC protein is rapidly degraded in vivo but is greatly stabilized through an interaction with UmuD′ [71,72]. The fact that the mutant R279C protein is stable in vivo suggests that it still retains the ability to interact with UmuD′. Thus, it is possible that R279 represents a region of UmuC that interacts with another protein normally required for translesion replication. We believe the most likely candidate is RecA protein, which if bound to single stranded DNA, would be predicted to be in contact with the back of the polymerase "hand".

Two other mutants are also located on the surface of the protein, but are likely to have a reduced ability to promote both spontaneous and damage induced

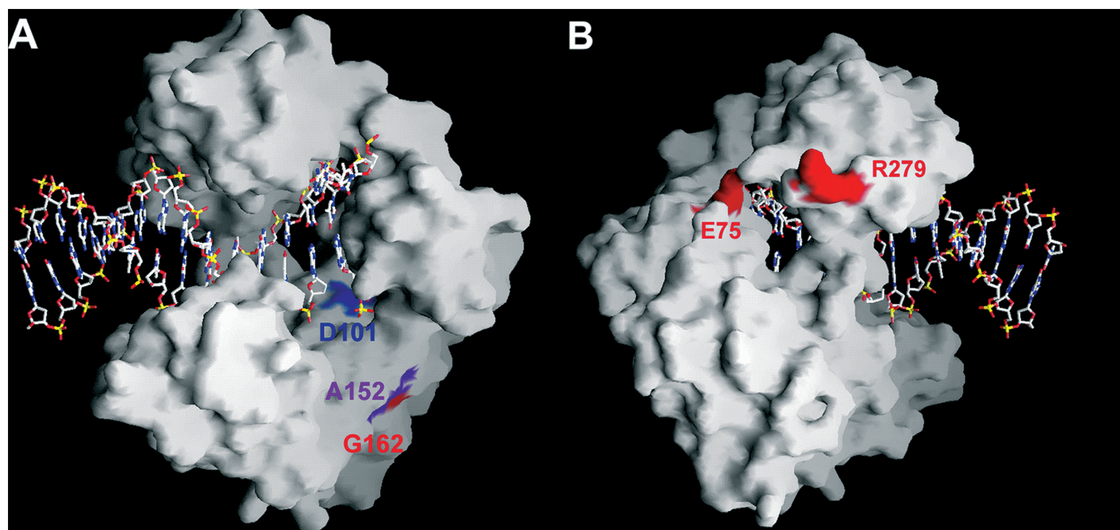

Fig. 3. Surface location of missense mutations in *E. coli* UmuC that results in reduced levels of SOS mutagenesis. Both views depict a molecular surface representation of Dpo4 and the location of the bound primer–template (stick representation). Residues are labeled based upon *E. coli* UmuC protein. (A) Front view showing DNA binding in the deep cleft between the thumb and little finger domains. UmuC104 protein substitutes D101 (in blue) for Asn at the catalytic active site of the polymerase. Although located on the surface of the protein, substitutions at A152 (pink) and G162 (red) are hypothesized to have structural ramifications rather then disrupt protein–protein interactions. (B) Rear view after a 180° rotation. UmuC36 protein substitutes E75 (red) with Lys. We hypothesize that this residue is involved in the interaction between UmuC and UmuD′. Substitution of R279 (red) with Cys greatly reduces the ability of UmuC to promote damage-induced mutagenesis. We hypothesize that this residue may be involved in an interaction with RecA protein. This figure was generated using GRASP [102].

mutagenesis because of structural defects, rather than in protein–protein interactions. The two residues in question are UmuC A152 and G162, which correspond to A158 and G162 in Dpo4, respectively (Fig. 3). Both residues are located in the "palm" domain, where there is only space to accommodate small side chain atoms. To accommodate a large substitution, conformational changes of the main chain in the surrounding residues from 158 to 163 have to occur, which we believe, will in turn, disturb the nearby incoming nucleotide binding site (including K159 in Dpo4; Fig. 2).

UmuC substitutions L133F and A140T gave similar phenotypes, in that they are defective for spontaneous mutagenesis, but moderately proficient for damage-induced mutagenesis [69]. These amino acid substitutions correspond to Dpo4 residues I138 and S145, respectively. Dpo4 I138 is located at the hydrophobic interface between the palm and fingers domain, while S145 is snugly surrounded by hydrophobic residues and makes a hydrogen bond important for maintaining the tertiary structure of the polymerase. Both I138 and S145 mutations would disturb protein stability or alter local structures in the polymerase, as in the case of the G162R mutation.

2.2. Mutations in E. coli UmuC that perturb its interaction with RecA

Although pol V is best characterized for its ability to promote damage induced mutagenesis, overproduction of UmuD′C also leads to an inhibition of homologous recombination in vivo [73–75]. Using a genetic screen for *E. coli* that were insensitive to the anti-recombinogenic actions of UmuD′C, Sommer et al., were able to isolate a number of mutant RecA proteins [75]. Sequence analysis of the mutant *recA* alleles revealed that many resulted in amino acids substitutions which are located on the surface of the RecA protein at, or near, the head–tail interface between RecA protomers in a nucleoprotein filament [75]. Based upon the location of these *recA* mutations, it was hypothesized that UmuD′C might bind to RecA

in two modes; one would be at the tip of a RecA nucleoprotein filament and would serve a dual role in targeting the Umu complex to sites of DNA damage while concomitantly capping the RecA nucleoprotein filament; the other would be in the deep helical groove of the RecA nucleoprotein filament itself, where it would act as a competitive inhibitor of double-stranded DNA-binding [75,76]. Recent electron microscopic studies of UmuD′C binding to a RecA nucleoprotein filament suggest that both hypotheses are, in fact, correct. At low concentrations, the UmuD′C complex binds preferentially to the tip of the RecA filament, but as the concentration of UmuD′C increases, it occupies the deep helical groove of the nucleoprotein filament [77].

Utilizing the same power of genetic selection used in the earlier studies, Sommer et al., recently isolated a number of suppressor mutations in UmuD′C that recover the ability to inhibit recombination even in the presence of a Umu-resistant (D112G) RecA protein [78]. Interestingly, although able to inhibit recombination when overproduced, many of the UmuC mutants exhibited a reduced ability to promote damage-induced mutagenesis, suggesting that translesion replication had been compromised at the expense of inhibiting RecA-mediated homologous recombination [78]. Alignment of the seven UmuC mutants isolated by Sommer et al., (UmuC F10L; Y270C; K277E; F287S/L; K342Q; F351I) to Dpo4 suggests that many may affect the ability of the polymerase to interact with DNA (Fig. 4). For example, UmuC K277 (Dpo4 K275) is on the surface of the protein and is located within the DNA-binding cleft of the polymerase. The wild-type Lys residue interacts with the backbone of the template DNA and substitution for Asp most likely reduces or abolishes these interactions. Similarly, F287 (Dpo4 H285) is on the surface of the protein and close to the template strand (Fig. 4). The F287S or F287L substitutions are therefore predicted to perturb DNA–protein interactions. UmuC K342 (Dpo4, K339), is also on the surface of the protein and interacts with the backbone of the primer (Fig. 4). UmuC F10, corresponds to Dpo4 F11. The side chain of F11 normally fits snugly in the hydrophobic core of the protein and is adjacent to the active site D105 residue (Fig. 2). Substitution of Phe with Leu would require repacking of the hydrophobic core and which will inevitably alter the conformation

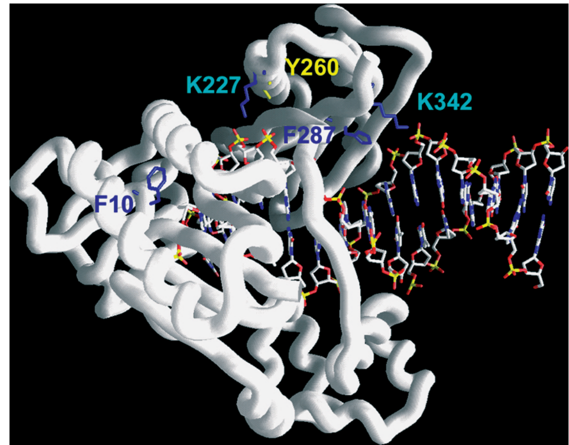

Fig. 4. Location of mutations within UmuC which result in a mutant protein unable to inhibit RecA-dependent homologous recombination. These mutants were isolated by Sommer et al. [78] and their location in UmuC mapped onto the structure of Dpo4. Residues are labeled based upon *E. coli* UmuC protein. Dpo4 is shown in a worm representation with the backbone in white and the side chains of F10, K277, F287 and K342 colored in blue and Y270 in yellow, in keeping with the color scheme used in Fig. 1. With the exception of F10, the remaining residues are located in the little finger domain of the polymerase. This figure was generated using GRASP [102].

of the active site of the polymerase. UmuC Y270 (Dpo4 A268 is located in the hydrophobic core of the little finger domain. Substitution of Y270 with Cys would probably require some adjustment of the neighboring residues so as to fill the void left by the bulky Tyr side chain. This rearrangement may destabilize the protein or alter conformations of the nearby surface residues, which in turn may change protein–DNA or protein–protein interactions.

Last but not least, is F351I (corresponding to Dpo4 K348). This residue is located in the very extreme C-terminal little finger domain of Dpo4 that remains unstructured, so the effects of this substitution are difficult to ascertain. Recent studies suggest, however, that this region is remarkably close to the proposed β-clamp binding motif of UmuC [79]. We hypothesize that the F351I substitution might therefore reduce the ability of pol V to interact with the β-clamp and subsequently remain engaged at the primer-terminus (Fig. 5).

We believe the location of all seven UmuC mutations reported by Sommer et al., are consistent

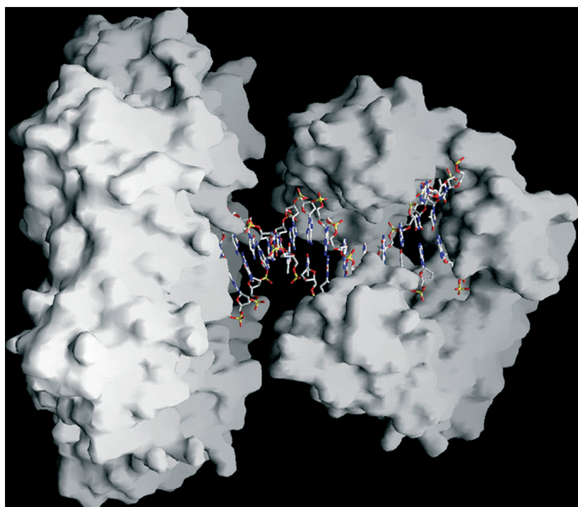

Fig. 5. Model of how a replicative sliding clamp and Y-family polymerase might appear at a primer terminus. The view depicts the surface representation of the *E. coli* β-clamp (2POL) behind Dpo4 (1JX4). It is assumed that the two proteins physically interact, but until such time as the residues involved in the clamp–Dpo4 interaction are identified, we have depicted a small gap between the two proteins. This model gives the reader a general idea of the relative sizes of the proteins and how they might be associated at a primer terminus. This figure was generated using GRASP [102].

with the hypothesis that they reduce the ability of the pol V complex to bind at the primer–template terminus and in doing so, reduce the ability of the polymerase to promote SOS mutagenesis. However, we believe that the mutant UmuC proteins retain their ability to bind to the deep helical groove of RecA, where they inhibit homologous recombination [78].

2.3. Mutations in E. coli DinB that result in a loss of polymerase activity and frameshift mutagenesis in vivo

To date, there have been no reports in the literature of mutant *dinB* alleles that have been isolated in a random screen for loss of cellular function. Rather, the reported *dinB* alleles have been generated by site-directed mutagenesis at D8, R49, D103 and E104 residues that are highly conserved in other members of the Y-family polymerases [24]. These residues correspond to D7, R51, D105 and E106 in Dpo4. As noted above, D7, D105 and E106 represent the active site of the enzyme, with D7 and D105 chelating the divalent metal ion required for catalysis (Fig. 2). As one might predict, substitution of the essential charged residue with alanine results in a dramatic decrease in frameshift mutagenesis in vivo [24]. R51 of Dpo4 is in close proximity to the active site and directly contacts the triphosphate moiety of an incoming nucleotide (Fig. 2). An R49F substitution in *E. coli* DinB not only reduces the extent of frameshift mutagenesis in vivo [24] but also adaptive mutagenesis [80].

2.4. A possible model for interactions between Dpo4 and a β/PCNA-like sliding clamp

It is becoming increasingly clear that like high-fidelity DNA polymerases, the catalytic activity of Y-family polymerases is stimulated via an interaction with a sliding clamp that helps tether the polymerase to DNA [25,53,54,63–67,81,82]. In *E. coli,* the clamp is the β subunit of pol III holoenzyme, while in eukaryotes it is the PCNA [83,84]. While neither share any significant primary amino acid homology and β is a dimer in solution while PCNA is a trimer, they both share similar quaternary structures, resembling a hexagon-like ring or “donut” that encircles DNA [85,86]. Recent studies suggest that each clamp may recognize and bind to structurally similar protein motifs [64,79,87–89]. The β-binding motif in *E. coli* DinB has been experimentally shown to include a stretch of hydrophobic amino acids consisting of Leu-Val-Leu-Gly-Leu, at the very C-terminus of the *E. coli* DinB protein [63]. *S. solfataricus* P2 [90] like other crenarchaea [91,92], apparently contains three PCNA homologs (Sso0397; Sso0405; Sso1047) and a potential PCNA binding site is located at the C-terminus of Dpo4 in an equivalent location to the β-binding site of the *E. coli* DinB protein (unpublished observations). An almost identical sequence is found in the related *S. solfataricus* P1 Dbh protein [14], which helps explain the results of Gruz et al., who observed stimulation of *S. solfataricus* P1 Dbh (pol Y1) activity in the presence of RFC and two of the three potential PCNA homologs from *S. solfataricus* P2 [53].

It seems reasonable to expect, therefore, that like other Y-family polymerases, the catalytic activity of

Dpo4 will be enhanced via an interaction with its cognate replicative clamp. The structure of the *S. solfataricus* PCNA-like clamp has yet to be determined, but it is likely to be somewhat similar to that reported for the *E. coli* β-clamp and PCNA proteins [85,86]. Although we do not know the exact location of the PCNA-binding site in Dpo4 and where it interacts with the PCNA-like clamp, we do know the relative locations/orientations of the two proteins on DNA; the polymerase will be at the primer-terminus and the clamp will encircle the double-stranded DNA behind the polymerase. As a consequence, we have generated a model for how one might envisage that the two proteins could associate on DNA (Fig. 5).

2.5. The Rev1-1 mutation in S. cerevisiae REV1

Given that the *REV1* locus was first identified in 1971 [1], it should, perhaps, be considered the founding member of the Y-family polymerases. Furthermore, the *S. cerevisiae* Rev1 protein was the first member of the superfamily shown to possess nucleotidyl transferase activity [20]. However, the fact that Rev1 only utilizes deoxycytosine tends to result in it getting less "exposure" than other members of the family, which utilize all four nucleoside triphosphates. Recent studies suggest that the mouse Rev1 protein can also incorporate dGMP and dTMP opposite template guanine, suggesting that Rev1-like proteins may be more "polymerase-like" than previously thought [93]. Unfortunately, the only mutant of *REV1* reported in the literature to date, that was isolated during a screen for loss of function, was the original Lemontt allele, *REV1-1* which results in a G193R substitution within Rev1 [4]. The location of the mutation is outside of the conserved polymerase "palm/thumb/finger/little finger" domains and consistent with this observation, is the fact that the Rev1-1 protein retains considerable deoxycytidyl transferase activity in vitro [94].

2.6. Mutations in S. cerevisiae RAD30 that result in a loss of function

The *RAD30* gene was identified and initially characterized at the genetic level nearly 5 years ago [21,22]. However, to date, there are no reports in the literature of missense mutations that have been identified in screens for a gain or loss of function in vivo. In an attempt to identify the catalytic site of pol η, Kondratick et al. utilized site-directed mutagenesis to generate alanine substitutions at a number of highly conserved acidic residues found in Rad30 and its orthologs [95]. The importance of these acidic amino acids is further highlighted by the fact that several of the mutated residues are also located within the five motifs previously identified in the UmuC/DinB/Rev1 proteins [14] and which we now know constitute the catalytic domain of the Y-family polymerases. Of the nine alanine mutants generated by Kondratick et al., three (D30A; D156A; and E157A) ultimately turned out to be at the active site of the enzyme (Figs. 2 and 6A) and are devoid of enzymatic activity [95]. Of the six other mutants, only E39A conferred a similar phenotype. Although E39 is not in the active site of the enzyme per se, it is close by and plays an important role in maintaining the tertiary structure of the finger domain of the polymerase (Fig. 6B). Alanine substitutions at E79, D160, D228, D235 and D239, had very little effect on the enzyme's ability to bypass a *cis–syn* thymine–thymine dimer in vitro [95]. Furthermore, both the E79A and D228A substitutions, when introduced into UV-sensitive *rad30Δ rad5Δ* strains, significantly increased cell viability, indicating that the mutants are also fully functional in vivo [95]. Both E79 and D228 are located on the surface of pol η (Fig. 6A) and although these amino acids are highly conserved in members of the Y-family, substitutions with alanine at these residues clearly has no significant impact on the activity of the enzyme in vitro or in vivo. In contrast, although apparently functional in vitro, D160A, D235A and D293A substitutions resulted in intermediate levels of UV-survival, suggesting that their in vivo activities are somewhat reduced [95]. D160 and D235 are located in the palm domain of the enzyme, while D293 is in the thumb domain (Fig. 6A). All three residues play important roles in maintaining the tertiary structure of the enzyme. Thus, we hypothesize that the reduced ability of the D160A, D235A and D293A mutant enzymes to restore UV-resistance in *rad30Δ rad5Δ* strains, may simply result from lower-steady state levels of the mutant polymerases, due to the rapid proteolysis of the unstable protein in vivo.

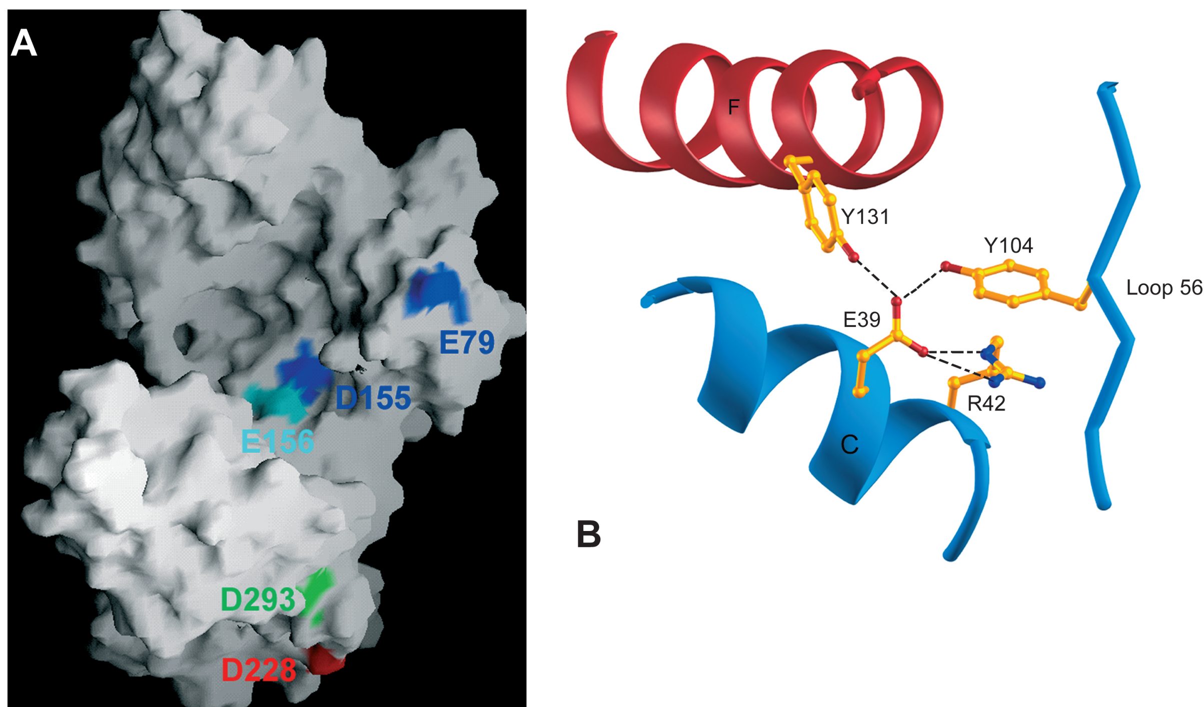

Fig. 6. Location of site-directed missense mutants of *S. cerevisiae* pol η that substitute a conserved carboxylate residue with alanine. These mutants were isolated by Kondratick et al. [95] and their location in based upon the catalytic core of pol η reported by Trincao et al. [40]. (A) Surface representation of pol η. D155 (blue) and E156 (turquoise) are at the active site of the enzyme. D30 is also at the active site but is buried and is not shown here (see Fig. 2 for comparison). E79 (blue), D228 (red) and D293 (green) are located on the surface of the protein, but substitution with alanine has little effect on the ability to bypass a thymine–thymine dimer in vitro. (B) Structural importance of E39. In this close up view, one can see that E39 makes hydrogen bonds (dash black line) with R42, Y104 and Y131. These helps maintain the structure of the finger domain (α-helix C), the loop between β-strands 5 and 6 (in blue) and the palm domain (α-helix F, in red). The (A) was generated using GRASP [102] and (B) using ribbons [101].

2.7. *Missense mutations in human DNA polymerase η that result in the xeroderma pigmentosum variant (XP-V) phenotype*

Most of the mutations identified in the human *POLH* gene that give rise to the XP-V phenotype result in a truncated protein [27,28,96,97]. In a recent study of 21 XP-V patients, Broughton et al. [97], were, however, successful in identifying mutations in patients that would be expected to produce full-length or close to full-length pol η protein. Four were missense mutations (R111H; T122P; G263V; and R361S), while the fifth was a small in-frame deletion that resulted in the loss of Leu75 codon [97]. The effect that these substitutions might have on human pol η has previously been discussed based upon the location of the mutation relative to Dpo4 [97]. In Fig. 7, we have mapped the mutations in human pol η on to the crystallized catalytic core of *S. cerevisiae* pol η [40]. In general, the location and the effect of the amino acid substitution remains the same whether the mutations are mapped to Dpo4 or *S. cerevisiae* pol η. Human L75 corresponds to *S. cerevisiae* L87 and is located on a β-strand within the hydrophobic core of the finger domain and deletion of this amino acid is likely to destabilize the tertiary structure of the entire protein. Arg111 (*S. cerevisiae*, R151) is in the palm domain and its aliphatic arm is involved in hydrophobic packing. The R111H substitution is likely to introduce a charged side chain into a hydrophobic environment and destabilize the protein. Thr 122 (*S. cerevisiae*, G162 and S112 of Dpo4) also plays

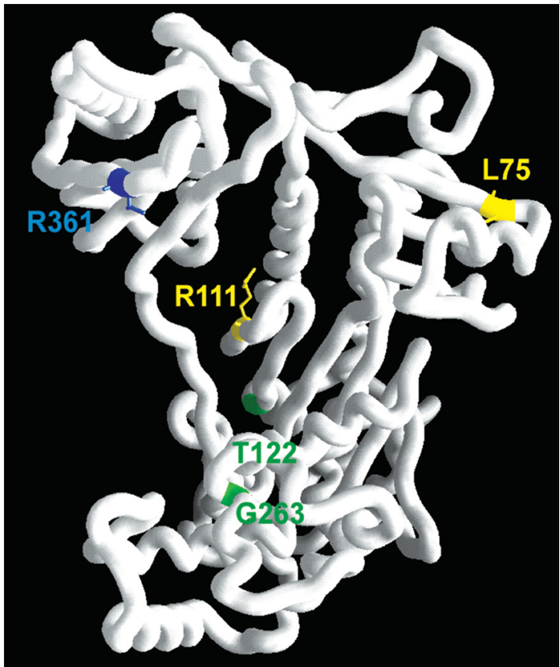

Fig. 7. Missense mutations in XP-V patients modeled onto *S. cerevisiae* pol η. The missense mutations in human pol η were identified by Broughton et al. [97] and are mapped onto the catalytic core of *S. cerevisiae* pol η [40]. The numbering is based upon the position of the mutation in human pol η. In this representation, pol η's Cα backbone is depicted as a worm. The worm is colored in yellow for L75 and R111 (hydrophobic residues); in green for T122 and G263 (structurally important); and blue for R361 (interacts with DNA) in accordance with their putative function (see Fig. 1 for color scheme).

an important role in maintaining the structure of the palm domain. Substitution of proline at the site will result in steric hindrance with a close by Asp residue (Asp160 of *S. cerevisiae* and Asp110 of Dpo4) and will, as a consequence, disrupt the palm domain. The G263V substitution is located near the N-terminus of helix K in the thumb domain, which contacts the template strand of DNA. G263 is conserved in pol η's in a variety of organisms because there is no space for any side chain atoms in the tertiary structures. Substitution of Val for Gly at amino acid 263 causes a clash with neighboring residues and would require structural changes from residues 253 to 264 to accommodate such a substitution. This, in turn, may alter the protein structure and protein–DNA interactions. Finally, R361S is in the little finger domain. In Dpo4 this residue corresponds to I295, while in *S. cerevisiae* it is V454. The side change of R361 points into the major groove of DNA and substitution or Arg with Ser will obviously effect the ability of the enzyme to bind DNA.

2.8. Missense mutations in the "fingers" domain of human DNA polymerase η

Very recently, Glick et al., utilized a yeast-based complementation assay to isolate several hundred biologically active mutants of human pol η [98]. In particular, they replaced the wild-type *POLH* sequence with a synthetic oligonucleotide containing approximately 8% random substitutions within a stretch of 66 nucleotides encoding human pol η Y52 to P73 [98]. This region encompasses the conserved motif II of the Y-family polymerases (Fig. 1) and is located in the fingers domain of the polymerase [39–42]. Biologically active mutants were selected based upon their ability to increase UV-resistance to normally UV-sensitive *rad30Δ rad52Δ* strains [21]. Using such an approach, Glick et al., isolated over 500 independent mutants. Sequence analysis of 210 of these mutants revealed that although motif II is conserved throughout evolution, many of the amino acids could be substituted without any apparent loss of function. The exceptions were Y52 and A54, which remained invariant. Both residues are located in the "substrate lid" of the finger domain. Y52 (equivalent to Y48 of Dpo4 and Y64 of *S. cerevisiae* pol η) stabilizes the protein structure and the incoming nucleotide by stacking between the conserved Arg residue 3 amino acids upstream (R51 of Dpo4 and R67 of pol η) and a conserved Pro residue in the palm domain (P160 of Dpo4 and P280 of *S. cerevisiae* pol η; Fig. 8). A54 is invariant because its side chain fits perfectly into the hydrophobic core of the finger domain. The absolute conservation of this residue indicates the importance of maintaining the exact conformation of this substrate lid in all Y-family polymerases for binding of the incoming nucleotide and polymerase catalysis (Fig. 8). Interestingly, although certain residues underwent multiple substitutions, others such as R55L, G58 and C72 (R51, G54 and L68, respectively in Dpo4; Figs. 1 and 8) could only be substituted in combination with other amino acids, suggesting that they too, play important roles in catalysis. Indeed, together with A54 and Y52, they define the

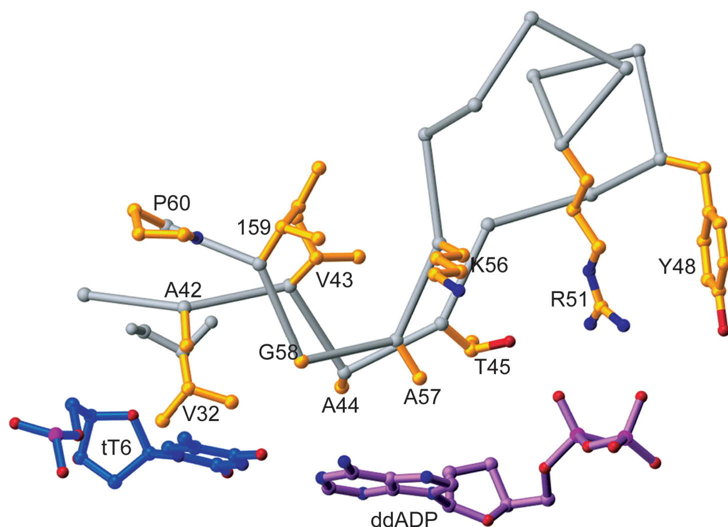

Fig. 8. Close up view of the fingers domain of Dpo4 encompassing the substrate lid. The incoming nucleotide (ddADP, pink) and its corresponding template (tT6, blue) are shown along with residues in the finger domain that constitute the substrate lid. The Cα backbone residue of V32 and A42 to P60 (42-AVATANYEARKFGVKAGIP-60) are colored gray and side chains are in yellow. These Dpo4 residues correspond to residues G46 to W64 (46-GIIAVSYEARAFGVTRSMW-64) in pol η. The ability of the enzyme to accommodate bulky DNA adducts in the active site of the enzyme appears to be determined in part, by residues in the substrate lid. For example, substitution of S62G in human pol η (corresponding to G58 in Dpo4), results in an enzyme that bypasses a variety of DNA lesions with greater efficiency than the wild-type polymerase [98]. This figure was generated using ribbons [101].

conformation of the substrate lid. For instance, R55 stacks with Y52 and interacts with the triphosphate moiety of the incoming nucleotide; G58 defines a rare structure (left handed turn) to position R55, while C72 interacts with the conserved F57 and defines the pockets for A54. In addition to identifying critical residues essential or important for catalysis, Glick et al., also identified one substitution, S62G, that not only enhanced the catalytic activity of the enzyme on undamaged DNA, but also its ability to bypass certain DNA lesions (e.g. a *cis–syn* thymine dimer; an abasic site; 8-oxo-guanine; O^4-methylthymine; and O^6-methyguanine) [98]. Human pol η S62 corresponds to *S. cerevisiae* pol η M74 and Dpo4 G58. In Dpo4, G58 is the residue that lies upon the template base (Fig. 8). Because of its small size, we believe that the Gly residue may allow better access of the bulky DNA adducts into the active site of the enzyme and therefore increases the efficiency of translesion replication. However, it is also clear from the ternary structure of Dpo4 [41] and the mutational studies of Glick et al. [98], that the overall topology of the substrate lid is, perhaps, the major determinant of lesion bypass. Dpo4, for example, naturally has Gly at residue 58 and while sharing some lesion bypass properties akin to pol η, it does not bypass a *cis–syn* thymine dimer with the same catalytic efficiency as pol η. It is interesting to note that certain residues in the substrate lid are quite variable within the Y-family polymerases. It seems reasonable to speculate that such variability may therefore provide the structural basis for their vastly different lesion bypass properties.

3. Concluding remarks

The aim of this review is to highlight how the structure of Y-family polymerases relates to their function. To achieve this goal, we mapped the location of several missense mutations on to the structure of Dpo4 or the catalytic core of *S. cerevisiae* pol η. Defects in some of the missense mutants could simply be attributed to the effect that the substitutions have on the overall stability of the enzyme rather than to a catalytic defect. For example, substitution of a polar residue with a charged residue in the hydrophobic core of the enzyme

would be structurally disastrous. Other substitutions are predicted to have little effect on the overall topology of the enzyme, yet have huge effects on regions of functional importance. Such regions clearly include the active site, the region surrounding the incoming nucleotide or the template base and its overall ability to bind DNA. In contrast, other substitutions were mapped to regions on the surface of the polymerase and therefore most likely represent regions where the polymerases interact with their protein partners.

Our study not only illuminates the structural basis for the phenotypes of many missense Y-family mutants; some of which were identified over 25 years ago [2,3]; others just a couple of months ago [97,98], but it also hopefully provides clues to the molecular mechanisms by which they promote translesion replication. However, the story is far from over. Now that the structures of several Y-family polymerases are in hand, the next step will be to generate mutations that address specific questions pertaining to their biochemical properties. For example, what is the structural basis for pol ι's preference to incorporate guanine opposite template thymine [29–31] or uracil [99], or why does Rev1 predominantly utilizes deoxycytosine over the other nucleoside triphosphates? We should also not forget that in addition to their conserved polymerase domain, many of the proteins contain significant C-terminal extensions/domains that clearly play an important role in the biological function of the respective Y-family polymerase [69,89,95,100]. Hopefully, in the not too distant future, tertiary structures of these C-terminal domains will be elucidated. Perhaps they will provide just as many clues and insights into possible protein–protein interactions and functions as those afforded to us by the crystal structures of the catalytic domains of the Y-family polymerases?

Acknowledgements

This work was supported by funds from the National Institutes of Health Intramural Research Program.

References

[1] J.F. Lemontt, Mutants of yeast defective in mutation induced by ultraviolet light, Genetics 68 (1971) 21–33.

[2] T. Kato, Y. Shinoura, Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutations by ultraviolet light, Mol. Gen. Genet. 156 (1977) 121–131.

[3] G. Steinborn, Uvm mutants of *Escherichia coli* K12 deficient in UV mutagenesis. Part I. Isolation of *uvm* mutants and their phenotypical characterization in DNA repair and mutagenesis, Mol. Gen. Genet. 165 (1978) 87–93.

[4] F.W. Larimer, J.R. Perry, A.A. Hardigree, The *REV1* gene of *Saccharomyces cerevisiae*: isolation, sequence and functional analysis, J. Bacteriol. 171 (1989) 230–237.

[5] Y. Kitagawa, E. Akaboshi, H. Shinagawa, T. Horii, H. Ogawa, T. Kato, Structural analysis of the *umu* operon required for inducible mutagenesis in *Escherichia coli*, Proc. Natl. Acad. Sci. U.S.A. 82 (1985) 4336–4340.

[6] K.L. Perry, S.J. Elledge, B. Mitchell, L. Marsh, G.C. Walker, *umuDC* and *mucAB* operons whose products are required for UV light and chemical-induced mutagenesis: UmuD, MucA and LexA products share homology, Proc. Natl. Acad. Sci. U.S.A. 82 (1985) 4331–4335.

[7] C.M. Smith, W.H. Koch, S.B. Franklin, P.L. Foster, T.A. Cebula, E. Eisenstadt, Sequence analysis of the *Salmonella typhimurium* LT2 *umuDC* operon, J. Bacteriol. 172 (1990) 4964–4978.

[8] S.M. Thomas, H.M. Crowne, S.C. Pidsley, S.G. Sedgwick, Structural characterization of the *Salmonella typhimurium* LT2 *umu* operon, J. Bacteriol. 172 (1990) 4979–4987.

[9] H. Ohmori, E. Hatada, Y. Qiao, M. Tsuji, R. Fukuda, *dinP*, a new gene in *Escherichia coli*, whose product shows similarities to UmuC and its homologues, Mutat. Res. 347 (1995) 1–7.

[10] T. Nohmi, A. Hakura, Y. Nakai, M. Watanabe, S.Y. Murayama, T. Sofuni, *Salmonella typhimurium* has two homologous but different *umuDC* operons: cloning of a new *umuDC*-like operon (*samAB*) present in a 60 MDa cryptic plasmid of *S. typhimurium*, J. Bacteriol. 173 (1991) 1051–1063.

[11] C. Ho, O.I. Kulaeva, A.S. Levine, R. Woodgate, A rapid method for cloning mutagenic DNA repair genes: isolation of *umu*-complementing genes from multidrug resistance plasmids R391, R446b and R471a, J. Bacteriol. 175 (1993) 5411–5419.

[12] O.I. Kulaeva, J.C. Wootton, A.S. Levine, R. Woodgate, Characterization of the *umu*-complementing operon from R391, J. Bacteriol. 177 (1995) 2737–2743.

[13] G.W. Sundin, S.P. Kidambi, M. Ullrich, C.L. Bender, Resistance to ultraviolet light in *Pseudomonas syringae*: sequence and functional analysis of the plasmid-encoded *rulAB* genes, Gene 177 (1996) 77–81.

[14] O.I. Kulaeva, E.V. Koonin, J.P. McDonald, S.K. Randall, N. Rabinovich, J.F. Connaughton, A.S. Levine, R. Woodgate, Identification of a DinB/UmuC homolog in the archeon *Sulfolobus solfataricus*, Mutat. Res. 357 (1996) 245–253.

[15] J.P. McDonald, V. Rapic-Otrin, J.A. Epstein, B.C. Broughton, X. Wang, A.R. Lehmann, D.J. Wolgemuth, R. Woodgate, Novel human and mouse homologs of *Saccharomyces cerevisiae* DNA polymerase η, Genomics 60 (1999) 20–30.

[16] V.L. Gerlach, L. Aravind, G. Gotway, R.A. Schultz, E.V. Koonin, E.C. Friedberg, Human and mouse homologs of

Escherichia coli DinB (DNA polymerase IV), members of the UmuC/DinB superfamily, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 11922–11927.

[17] H. Ohmori, E.C. Friedberg, R.P.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.

[18] B.A. Bridges, R. Woodgate, The two-step model of bacterial UV mutagenesis, Mutat. Res. 150 (1985) 133–139.

[19] E.C. Friedberg, G.C. Walker, W. Siede, DNA Repair and Mutagenesis, American Society for Microbiology, Washington, DC, 1995.

[20] J.R. Nelson, C.W. Lawrence, D.C. Hinkle, Deoxycytidyl transferase activity of yeast Rev1 protein, Nature 382 (1996) 729–731.

[21] J.P. McDonald, A.S. Levine, R. Woodgate, The *Saccharomyces cerevisiae RAD30* gene, a homologue of *Escherichia coli dinB* and *umuC*, is DNA damage inducible and functions in a novel error-free post-replication repair mechanism, Genetics 147 (1997) 1557–1568.

[22] A.A. Roush, M. Suarez, E.C. Friedberg, M. Radman, W. Siede, Deletion of the *Saccharomyces cerevisiae* gene *RAD30* encoding an *Escherichia coli* DinB homolog confers UV radiation sensitivity and altered mutability, Mol. Gen. Genet. 257 (1998) 686–692.

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

[24] J. Wagner, P. Gruz, S.R. Kim, M. Yamada, K. Matsui, R.P.P. Fuchs, T. Nohmi, The *dinB* gene encodes an novel *Escherichia coli* DNA polymerase (DNA pol IV) involved in mutagenesis, Mol. Cell 4 (1999) 281–286.

[25] M. Tang, X. Shen, E.G. Frank, M. O'Donnell, R. Woodgate, M.F. Goodman, $UmuD'_2C$ is an error-prone DNA polymerase, *Escherichia coli* DNA pol V, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 8919–8924.

[26] N.B. Reuven, G. Arad, A. Maor-Shoshani, Z. Livneh, The mutagenesis protein UmuC is a DNA polymerase activated by UmuD′ RecA and SSB and is specialized for translesion replication, J. Biol. Chem. 274 (1999) 31763–31766.

[27] 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 η, Nature 399 (1999) 700–704.

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

[29] A. Tissier, J.P. McDonald, E.G. Frank, R. Woodgate, pol ι, a remarkably error-prone human DNA polymerase, Genes Dev. 14 (2000) 1642–1650.

[30] R.E. Johnson, M.T. Washington, L. Haracska, S. Prakash, L. Prakash, Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions, Nature 406 (2000) 1015–1019.

[31] Y. Zhang, F. Yuan, X. Wu, Z. Wang, Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase ι, Mol. Cell. Biol. 20 (2000) 7099–7108.

[32] R.E. Johnson, S. Prakash, L. Prakash, The human *DINB1* gene encodes the DNA polymerase pol θ, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 3838–3843.

[33] E. Ohashi, T. Ogi, R. Kusumoto, S. Iwai, C. Masutani, F. Hanaoka, H. Ohmori, Error-pone bypass of certain DNA lesions by the human DNA polymerase κ, Genes Dev. 14 (2000) 1589–1594.

[34] 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 κ in vitro, Nucleic Acids Res. 28 (2000) 4138–4146.

[35] V.L. Gerlach, W.J. Feaver, P.L. Fischhaber, E.C. Friedberg, Purification and characterization of pol κ, a DNA polymerase encoded by the human *DINB1* gene, J. Biol. Chem. 276 (2001) 92–98.

[36] J. Ito, D.K. Braithwaite, Compilation and alignment of DNA polymerase sequences, Nucleic Acids Res. 19 (1991) 4045–4057.

[37] D.K. Braithwaite, J. Ito, Compilation, alignment and phylogenetic relationships of DNA polymerases, Nucleic Acids Res. 21 (1993) 787–802.

[38] I.K. Cann, Y. Ishino, Archaeal DNA replication: identifying the pieces to solve a puzzle, Genetics 152 (1999) 1249–1267.

[39] B. 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.

[40] 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 η implications for translesion DNA synthesis, Mol. Cell 8 (2001) 417–426.

[41] 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.

[42] L.F. Silvian, E.A. Toth, P. Pham, M.F. Goodman, T. Ellenberger, Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*, Nat. Struct. Biol. 8 (2001) 984–989.

[43] S. Doublie, S. Tabor, A.M. Long, C.C. Richardson, T. Ellenberger, Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution, Nature 391 (1998) 251–258.

[44] M.C. Franklin, J. Wang, T.A. Steitz, Structure of the replicating complex of a pol α family DNA polymerase, Cell 105 (2001) 657–667.

[45] H. Huang, R. Chopra, G.L. Verdine, S.C. Harrison, Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance, Science 282 (1998) 1669–1675.

[46] J.R. Kiefer, C. Mao, J.C. Braman, L.S. Beese, Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal, Nature 391 (1998) 304–307.

[47] Y. Li, S. Korolev, G. Waksman, Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation, EMBO J. 17 (1998) 7514–7525.

[48] T.A. Steitz, DNA polymerases: structural diversity and common mechanisms, J. Biol. Chem. 274 (1999) 17395–17398.

[49] T. Matsuda, K. Bebenek, C. Masutani, F. Hanaoka, T.A. Kunkel, Low fidelity DNA synthesis by human DNA polymerase η, Nature 404 (2000) 1011–1013.

[50] K. Bebenek, T. Matsuda, C. Masutani, F. Hanaoka, T.A. Kunkel, Proofreading of DNA polymerase η-dependent replication errors, J. Biol. Chem. 276 (2001) 2317–2320.

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

[52] F. Boudsocq, S. Iwai, F. Hanaoka, R. Woodgate, *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DNA polymerase with lesion-bypass properties akin to eukaryotic pol η, Nucleic Acids Res. 29 (2001) 4607–4616.

[53] P. Gruz, F.M. Pisani, M. Shimizu, M. Yamada, I. Hayashi, K. Morikawa, T. Nohmi, Synthetic activity of *Sso* DNA polymerase Y1, an archaeal DinB-like DNA polymerase, is stimulated by processivity factors proliferating cell nuclear antigen and replication factor C, J. Biol. Chem. 276 (2001) 47394–47401.

[54] M. Tang, P. Pham, X. Shen, J-S. Taylor, M. O'Donnell, R. Woodgate, M. Goodman, Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis, Nature 404 (2000) 1014–1018.

[55] J. Wagner, T. Nohmi, *Escherichia coli* DNA polymerase IV mutator activity: genetic requirements and mutational specificity, J. Bacteriol. 182 (2000) 4587–4595.

[56] R. Napolitano, R. Janel-Bintz, J. Wagner, R.P. Fuchs, All three SOS-inducible DNA polymerases (pols II, IV and V) are involved in induced mutagenesis, EMBO J. 19 (2000) 6259–6265.

[57] 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 κ, the product of the human *DINB1* gene, J. Biol. Chem. 275 (2000) 39684–39786.

[58] A. Vaisman, A. Tissier, E.G. Frank, M.F. Goodman, R. Woodgate, Human DNA polymerase ι promiscuous mismatch extension, J. Biol. Chem. 276 (2001) 30615–30622.

[59] R. Woodgate, M. Rajagopalan, C. Lu, H. Echols, UmuC mutagenesis protein of *Escherichia coli*: purification and interaction with UmuD and UmuD′, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 7301–7305.

[60] I. Bruck, R. Woodgate, K. McEntee, M.F. Goodman, Purification of a soluble UmuD′C complex from *Escherichia coli*: cooperative binding of UmuD′C to single-stranded DNA, J. Biol. Chem. 271 (1996) 10767–10774.

[61] P. Pham, J.G. Bertram, M. O'Donnell, R. Woodgate, M.F. Goodman, A model for SOS-lesion targeted mutations in *E. coli* involving pol V, RecA, SSB and β sliding clamp, Nature 409 (2001) 366–370.

[62] N.B. Reuven, G. Arad, A.Z. Stasiak, A. Stasiak, Z. Livneh, Lesion bypass by the *Escherichia coli* DNA polymerase V requires assembly of a RecA nucleoprotein filament, J. Biol. Chem. 276 (2001) 5511–5517.

[63] N. Lenne-Samuel, J. Wagner, H. Etienne, R.P. Fuchs, The processivity factor β controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis in vivo, EMBO Rep. 3 (2002) 45–49.

[64] L. Haracska, R.E. Johnson, I. Unk, B. Phillips, J. Hurwitz, L. Prakash, S. Prakash, Physical and functional interactions of human DNA polymerase η with PCNA, Mol. Cell. Biol. 21 (2001) 7199–7206.

[65] L. Haracska, C.M. Kondratick, I. Unk, S. Prakash, L. Prakash, Interaction with PCNA is essential for yeast DNA polymerase η function, Mol. Cell 8 (2001) 407–415.

[66] L. Haracska, R.E. Johnson, I. Unk, B.B. Phillips, J. Hurwitz, L. Prakash, S. Prakash, Targeting of human DNA polymerase ι to the replication machinery via interaction with PCNA, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 14256–14261.

[67] L. Haracska, I. Unk, R.E. Johnson, B.B. Phillips, J. Hurwitz, L. Prakash, S. Prakash, Stimulation of DNA synthesis activity of human DNA polymerase κ by PCNA, Mol. Cell. Biol. 22 (2002) 784–791.

[68] W.H. Koch, D.G. Ennis, A.S. Levine, R. Woodgate, *Escherichia coli umuDC* mutants: DNA sequence alterations and UmuD cleavage, Mol. Gen. Genet. 233 (1992) 443–448.

[69] R. Woodgate, M. Singh, O.I. Kulaeva, E.G. Frank, A.S. Levine, W.H. Koch, Isolation and characterization of novel plasmid-encoded *umuC* mutants, J. Bacteriol. 176 (1994) 5011–5021.

[70] H. Bates, B.A. Bridges, R. Woodgate, Mutagenic DNA repair in *Escherichia coli*. Part XX. Overproduction of UmuD′ protein results in suppression of the *umuC36* mutation in excision defective bacteria, Mutat. Res. 250 (1991) 199–204.

[71] E.G. Frank, M. Gonzalez, D.G. Ennis, A.S. Levine, R. Woodgate, In vivo stability of the Umu mutagenesis proteins: a major role for RecA, J. Bacteriol. 178 (1996) 3550–3556.

[72] E.G. Frank, D.G. Ennis, M. Gonzalez, A.S. Levine, R. Woodgate, Regulation of SOS mutagenesis by proteolysis, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 10291–10296.

[73] S. Sommer, A. Bailone, R. Devoret, The appearance of the UmuD′C protein complex in *Escherichia coli* switches repair from homologous recombination to SOS mutagenesis, Mol. Microbiol. 10 (1993) 963–971.

[74] F. Boudsocq, M. Campbell, R. Devoret, A. Bailone, Quantitation of the inhibition of Hfr × F^- recombination by the mutagenesis complex UmuD′C, J. Mol. Biol. 270 (1997) 201–211.

[75] S. Sommer, F. Boudsocq, R. Devoret, A. Bailone, Specific RecA amino acid changes affect RecA–UmuD′C interaction, Mol. Microbiol. 28 (1998) 281–291.

[76] W.M. Rehrauer, I. Bruck, R. Woodgate, M.F. Goodman, S.C. Kowalczykowski, Modulation of recombination function by the mutagenic UmuD′C protein complex, J. Biol. Chem. 273 (1998) 32384–32387.

[77] E.G. Frank, N. Cheng, C.C. Do, M.E. Cerritelli, I. Bruck, M.F. Goodman, E.H. Egelman, R. Woodgate, A.C. Steven, Visualization of two binding sites for the *Escherichia coli* $UmuD'_2C$ complex (DNA pol V) on RecA-ssDNA filaments, J. Mol. Biol. 297 (2000) 585–597.

[78] S. Sommer, G. Coste, A. Bailone, Specific amino acid changes enhance the anti-recombination activity of the UmuD′C complex, Mol. Microbiol. 35 (2000) 1443–1453.

[79] B.P. Dalrymple, K. Kongsuwan, G. Wijffels, N.E. Dixon, P.A. Jennings, A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 11627–11632.

[80] G.J. McKenzie, P.L. Lee, M.J. Lombardo, P.J. Hastings, S.M. Rosenberg, SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification, Mol. Cell 7 (2001) 571–579.

[81] M. Tang, I. Bruck, R. Eritja, J. Turner, E.G. Frank, R. Woodgate, M. O'Donnell, M.F. Goodman, Biochemical basis of SOS-induced mutagenesis in *Escherichia coli*: reconstitution of in vitro lesion bypass dependent on the $UmuD'_2C$ mutagenic complex and RecA, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 9755–9760.

[82] J. Wagner, S. Fujii, P. Gruz, T. Nohmi, R.P. Fuchs, The β-clamp targets DNA polymerase IV to DNA and strongly increases its processivity, EMBO Rep. 1 (2001) 484–488.

[83] A. Kornberg, T.A. Baker, DNA replication, Freeman, New York, 1992.

[84] I. Bruck, M. O'Donnell, The ring-type polymerase sliding clamp family, Genome Biol. 2 (2001) (Reviews 3001).

[85] X.P. Kong, R. Onrust, M. O'Donnell, J. Kuriyan, Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp, Cell 69 (1992) 425–437.

[86] T.S. Krishna, X.P. Kong, S. Gary, P.M. Burgers, J. Kuriyan, Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA, Cell 79 (1994) 1233–1243.

[87] D. Jeruzalmi, M. O'Donnell, J. Kuriyan, Crystal structure of the processivity clamp loader γ complex of *E. coli* DNA polymerase III, Cell 106 (2001) 429–441.

[88] M. O'Donnell, D. Jeruzalmi, J. Kuriyan, Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC, Curr. Biol. 11 (2001) 935–946.

[89] P. Kannouche, B.C. Broughton, M. Volker, F. Hanaoka, L.H. Mullenders, A.R. Lehmann, Domain structure, localization and function of DNA polymerase η, defective in xeroderma pigmentosum variant cells, Genes Dev. 15 (2001) 158–172.

[90] Q. She, R.K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M.J. Awayez, C.C. Chan-Weiher, I.G. Clausen, B.A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P.M. Gordon, I. Heikamp-De Jong, A.C. Jeffries, C.J. Kozera, N. Medina, X. Peng, H.P. Thi-Ngoc, P. Redder, M.E. Schenk, C. Theriault, N. Tolstrup, R.L. Charlebois, W.F. Doolittle, M. Duguet, T. Gaasterland, R.A. Garrett, M.A. Ragan, C.W. Sensen, J. Van Der Oost, The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 7835–7840.

[91] Y. Kawarabayasi, Y. Hino, H. Horikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, H. Kikuchi, Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7, DNA Res. 8 (2001) 123–140.

[92] K. Daimon, Y. Kawarabayasi, H. Kikuchi, Y. Sako, Y. Ishino, Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon *Aeropyrum pernix*: interactions with the two DNA polymerases, J. Bacteriol. 184 (2002) 687–694.

[93] Y. Masuda, M. Takahashi, S. Fukuda, M. Sumii, K. Kamiya, Mechanisms of dCMP transferase reactions catalyzed by mouse Rev1 protein, J. Biol. Chem. 277 (2002) 3040–3046.

[94] J.R. Nelson, P.E. Gibbs, A.M. Nowicka, D.C. Hinkle, C.W. Lawrence, Evidence for a second function for *Saccharomyces cerevisiae* Rev1 protein, Mol. Microbiol. 37 (2000) 549–554.

[95] C.M. Kondratick, M.T. Washington, S. Prakash, L. Prakash, Acidic residues critical for the activity and biological function of yeast DNA polymerase η, Mol. Cell. Biol. 21 (2001) 2018–2025.

[96] M. Yuasa, C. Masutani, T. Eki, F. Hanaoka, Genomic structure, chromosomal localization and identification of mutations in the xeroderma pigmentosum variant (*XPV*) gene, Oncogene 19 (2000) 4721–4728.

[97] B.C. Broughton, A. Cordonnier, W.J. Kleijer, N.G. Jaspers, H. Fawcett, A. Raams, V.H. Garritsen, A. Stary, M.F. Avril, F. Boudsocq, C. Masutani, F. Hanaoka, R.P. Fuchs, A. Sarasin, A.R. Lehmann, Molecular analysis of mutations in DNA polymerase η in xeroderma pigmentosum variant patients, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 815–820.

[98] E. Glick, K.L. Vigna, L.A. Loeb, Mutations in human DNA polymerase η motif II alter bypass of DNA lesions, EMBO J. 20 (2001) 7303–7312.

[99] A. Vaisman, R. Woodgate, Unique misinsertion specificity of pol ι may decrease the mutagenic potential of deaminated cytosines, EMBO J. 20 (2001) 6520–6529.

[100] M.D. Sutton, G.C. Walker, *umuDC*-mediated cold sensitivity is a manifestation of functions of the $UmuD_2C$ complex involved in a DNA damage checkpoint control, J. Bacteriol. 183 (2001) 1215–1224.

[101] M. Carson, Ribbon model of macromolecules, J. Mol. Graphics 5 (1987) 103–106.

[102] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, Proteins 11 (1991) 281–296.