# Commentary

# The many faces of DNA polymerases: Strategies for mutagenesis and for mutational avoidance

**Errol C. Friedberg\*, William J. Feaver, and Valerie L. Gerlach**

Laboratory of Molecular Pathology, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390-9072

The past year or so has witnessed the emergence of a plethora of prokaryotic and eukaryotic genes that are known or predicted to encode previously unidentified DNA polymerases (1–13). Some of these are members of an extended superfamily of prokaryotic and eukaryotic proteins called the UmuC/DinB nucleotidyl transferase superfamily, named after early discovered prokaryotic members (14, 15). This superfamily is represented presently by the UmuC, DinB, Rad30, and Rev1 subfamilies (14, 15). More recently identified polymerases resemble DNA replicative enzymes, and others seem to be related to the Pol $\beta$ and terminal transferase proteins.

A current working hypothesis is that when highly processive semiconservative DNA replication is arrested at lesions in DNA, the replicative machinery is displaced from the replication fork and replaced by these DNA polymerases. When the offending lesion has been bypassed successfully, the polymerase displacement/replacement process is reversed, and the replication machinery continues high-fidelity, highly processive DNA synthesis. There is substantial evidence that some of these prokaryotic polymerases are involved in such replicative bypass (translesion synthesis or TLS) of damaged DNA. Hence, the mechanism or mechanisms by which these enzymes function predictably lie at the heart of some types of DNA damage-induced mutagenesis in organisms such as *Escherichia coli*. A primary example is UmuC protein, one of the members of the UmuC subfamily. This protein complexes with UmuD′ protein (a proteolytically processed form of UmuD) to form a UmuD′$_2$C complex that facilitates TLS in the presence of *E. coli* DNA polymerase III holoenzyme (1–5, 13). The UmuD′$_2$C complex alone (most likely the UmuC polypeptide) is now called DNA polymerase V of *E. coli* (16–18). More recent studies have shown that DNA polymerase V bypasses thymine–thymine dimers, [6-4] photoproducts, and abasic sites in a highly error-prone manner *in vitro* (19). Additionally, this enzyme has poor fidelity on undamaged DNA primer-

templates, with error rates of $10^{-3}$ to $10^{-4}$ (19), substantially higher than those of most replicative DNA polymerases *in vitro* (13).

In addition to DNA polymerase V, a polymerase called DNA polymerase IV of *E. coli*, the product of the *dinB* gene and a member of the DinB subfamily, was purified recently (20). DNA polymerase IV is devoid of $3' \rightarrow 5'$ exonuclease activity and is strictly distributive in nature. Significantly, with respect to its role in spontaneous mutagenesis *in vivo*, the enzyme introduces frameshifts on misaligned primer-template substrates, resulting in −1 frameshift mutations (20). Most recently, it has been shown that a purified maltose-binding protein–DinB fusion protein is unable to bypass *cis-syn* thymine–thymine dimers, [6-4] photoproducts, or abasic sites in template DNA *in vitro* (19). Clearly, unlike DNA polymerase V, *E. coli* DNA polymerase IV is not able to support TLS at sites of chemically altered bases but may be able to support DNA synthesis at misaligned replication forks.

The spotlight is now shifting to the functions of some of the eukaryotic genes and their polypeptide products, in particular the biochemical demonstration that they are indeed DNA polymerases, as well as a consideration of their fidelity, processivity, and ability to support various types of TLS. A consistent theme that is now beginning to emerge is that a number of these DNA polymerases have poor fidelity on normal DNA primer-templates. Interestingly, however, this property may be inconsequential most of the time in living cells, because these polymerases normally may not have access to undamaged DNA. Indeed, the property of limited fidelity may be essential for the primary (and possibly exclusive) biological function of some of these enzymes, which is to negotiate noninstructional types of template damage or conformational distortions to allow normal replication to continue. Furthermore, these enzymes can support this function with relative accuracy or inaccuracy and hence generate mutations or not at such noninstruc-

tional sites, depending on the type of lesion and the particular polymerase in question. However, it remains to be proven that the poor replicational fidelity on undamaged DNA of some of these DNA polymerases that are already partially characterized and possibly of others that remain to be characterized is not biologically consequential.

Eukaryotic orthologs of the *E. coli dinB* gene have been identified in the genomes of the fission yeast *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, mice, and humans (14, 15, 21). The budding yeast *Saccharomyces cerevisiae* lacks this gene, and there is no obvious *dinB* ortholog in the genome of *Drosophila melanogaster*, the sequence of which was completed recently. The human *DINB1* and mouse *Dinb1* cDNAs have been cloned and partially characterized (14, 21). In a recent issue of PNAS, Johnson *et al.* (22) have shown that the polypeptide product of the human *DINB1* gene (which has been named pol $\kappa$ or pol $\theta$ by different groups; see below) is a DNA polymerase with properties very similar to those of the orthologous enzyme from *E. coli*. Johnson *et al.* (22) purified a glutathione *S*-transferase (GST)–DinB1 fusion protein after overexpression in *S. cerevisiae* and demonstrated that polymerase activity was inactivated after mutation of highly conserved residues. Like the *E. coli* enzyme, GST–human DinB1 fusion protein can catalyze extension of a misaligned primer template to generate −1 frameshifts. Additionally, like its *E. coli* ortholog, human GST–DinB1 fusion protein is unable to replicate past *cis-syn* thymine–thymine dimers, [6-4] photoproducts, or abasic sites. Remarkably, the frequency of misincorporation at single nucleotide sites in undamaged primer-template DNA by

**Table 1. Human DNA-directed DNA polymerases**

| Gene symbol | Name | Literature aliases | Chromosomal location | DNA polymerase family (refs. 41 and 42) | Function |
|---|---|---|---|---|---|
| *POLA* | pol α | | Xp22.1-p21.3 | B | Primer function in DNA replication |
| *POLB* | pol β | | 8p12-11 | X | Base excision repair |
| *POLG* | pol γ | | 15q24 | A | Mitochondrial DNA replication and repair |
| *POLG2* | pol γ2 | | 17q | | |
| *POLD1* | pol δ1 | | 19q13.3 | B | DNA replication; nucleotide and base excision repair |
| *POLD2* | pol δ2 | | 7 | | |
| *POLE* | pol ε | | 12q24.3 | B | DNA replication; nucleotide and base excision repair |
| *POLE2* | pol ε2 | | 14q21-22 | | |
| *POLZ* | pol ζ | *REV3L* | 6q21-22.2 | B | ?TLS |
| *POLH* | pol η | *RAD30A* *XPV* | ? | UmuC/DinB | Accurate TLS |
| *POLQ* | pol θ | *POLH* | 3q13.3 | A | ?DNA repair of crosslinks |
| *POLI* | pol ι | *RAD30B* | 18q21.1 | UmuC/DinB | ? |
| *POLK* | pol κ | *DINB1* *POLQ* | 5q13 | UmuC/DinB | ? |
| *POLL* | pol λ | | 10 | X | ?Meiosis-associated DNA repair |
| *POLM* | pol μ | | 7p13 | X | ?Somatic hypermutation |
| *REV1* | Deoxycytidyl transferase | | 2q11 | UmuC/DinB | TLS |

The names of genes and proteins are listed according to the guidelines of the Human Genome Organization Gene Nomenclature Committee and refs. 39 and 40.

the human GST–DinB1 fusion protein is $10^{-4}$ to $3 \times 10^{-3}$ (22).

Verification of the DNA polymerase activity of human DinB1 protein has been provided independently by (at least) two other laboratories. We have purified full-length GST–human DinB1 fusion protein to apparent homogeneity and have demonstrated DNA polymerase activity *in vitro* (W.J.F., V.L.G., and E.C.F., unpublished observations). Additionally, Ohmori and colleagues have purified a histidine-tagged truncated form of human DinB1 protein from baculovirus-infected insect cells (H. Ohmori, E. Ohashi, T. Ogi, R. Kusumoto, S. Iwai, C. Masutani, and F. Hanaoka, personal communication). The truncated protein is deleted of the C-terminal 312 amino acids, which includes two C2HC zinc cluster domains (14), but is nonetheless an active DNA polymerase. As reported by Johnson *et al.* (22) with the full-length DinB1–GST fusion protein, the truncated protein is unable to replicate past thymine–thymine dimers or [6-4] photoproducts at enzyme concentrations varying over two orders of magnitude. However, it is able to bypass abasic sites at high enzyme concentrations. This observation underscores one of the major problems associated with the evaluation of all DNA polymerases in reductionist *in vitro* systems, i.e., their fidelity on both undamaged and deliberately damaged substrates can be influenced significantly by enzyme concentration, template sequence context, and reaction conditions.

The Rad30 branch of the UmuC/DinB superfamily includes the protein products of the *S. cerevisiae* RAD30 gene and the orthologous human *POLH* (*RAD30A/XPV*) gene, which encode DNA polymerases called pol η (refs. 23–28; Table 1). In contrast to *E. coli* DNA polymerases V and IV and human DinB1 protein, human pol η (23) and its highly conserved yeast ortholog (25, 28) can replicate across *cis-syn* thymine–thymine dimers in template DNA, inserting adenines opposite the dimerized thymines. The processivity and fidelity of yeast pol η on undamaged DNA templates and on templates containing a single *cis-syn* thymine–thymine dimer have been examined recently in greater detail (29, 30). Yeast pol η has extremely limited processivity on undamaged primer-templates, incorporating only a few nucleotides during each enzyme–substrate encounter (29, 30). It is also a highly inaccurate enzyme, replicating undamaged DNA with a misincorporation frequency of $\approx 10^{-2}$ to $10^{-3}$, as measured at single nucleotide sites (29, 30).

Yeast pol η manifests the same limited processivity and high misincorporation frequency opposite a *cis-syn* thymine–thymine dimer as it does on undamaged DNA (30). Nonetheless, Washington *et al.* (30) point out that, if the average processivity of the enzyme is four nucleotides *in vivo*, these properties of the enzyme are sufficient to facilitate accurate bypass of thymine–thymine dimers >96% of the time. From a biological standpoint, this level of fidelity presumably translates into

a mechanism for most often negotiating such lesions accurately at replication forks. Like the yeast enzyme, human pol η is also extremely inaccurate when copying undamaged DNA. A GST–human pol η fusion protein has been shown to misincorporate nucleotides with a frequency of $10^{-2}$ to $10^{-3}$ (31). A recent independent study (32) monitored the average fidelity of purified histidine-tagged recombinant human pol η protein by measuring the frequency of all possible 12 mispaired bases generated *in vitro*. Depending on the particular mismatch measured, pol η inserted an incorrect nucleotide for every 18–380 nucleotides synthesized. As is the case with the yeast enzyme (30), human GST–pol η fusion protein copies *cis-syn* thymine–thymine dimers with the same limited fidelity with which it copies thymine monomers in undamaged DNA (31).

Hence, in contrast to an enzyme such as *E. coli* DNA polymerase V, which supports TLS in an inaccurate manner, the biological function of yeast and human pol η seems to be most obviously related to the avoidance of errors opposite *cis-syn* thymine–thymine dimers by catalyzing the incorporation of adenines. The fact of its poor fidelity on undamaged primer-templates *in vitro* may be irrelevant to its biological function *in vivo*, as evidenced by the fact that humans endowed with this enzyme activity are much more highly protected from the mutagenic and carcinogenic effects of sunlight than are those without it (24, 25). However, it remains to be determined whether human pol η is in fact "reading" dimerized thymine residues in template DNA or is simply following the so-called "A rule" by which some DNA polymerases insert adenine residues as the default nucleotide opposite noninstructional lesions. It is also clearly important to determine whether human and yeast pol η can support TLS across cytosine–thymine or cytosine–cytosine dimers and, if so, which nucleotides are incorporated most frequently in these situations. Given these caveats and the fact that pol η is a highly inaccurate DNA polymerase in general, it is misleading to refer to any type of DNA synthesis supported by this enzyme as error-free. Indeed, until we have a better understanding of the biological function or functions of all of the DNA polymerases addressed in this commentary, it may be prudent to refrain from using the terms "error-prone" and "error-free" at all.

Similar considerations may apply to *E. coli* DNA polymerase IV and its human ortholog DinB1 protein. These DNA polymerases may be specifically required to negotiate sites of stalled or arrested DNA replication precipitated by replicational slippage in a manner that frequently introduces mutations. As with pol η, the poor fidelity of DinB1 protein on normal

DNA may be biologically irrelevant unless the protein is overexpressed. It has in fact been demonstrated that overexpression of the *E. coli dinB* gene results in a large increase in the frequency of untargeted (spontaneous) mutations in plasmid DNA (reviewed in ref. 3). Similarly, overexpression of the orthologous mouse *dinb1* gene in cultured mouse cells results in an almost 10-fold increase in the level of spontaneous mutations (21). It remains an intriguing question as to whether the molecular mechanism of mutagenesis associated with overexpression of DinB1 protein is identical to that associated with normal levels of expression. Regardless, it is reasonable to expect that both expression of the human *DINB1* and mouse *Dinb1* genes and the stability of the polypeptide gene products are tightly regulated.

Additional DNA polymerases have recently been discovered. A paralog of the human *POLH* gene called *POLI* (*RAD30B*) has been shown to encode a protein with DNA polymerase activity designated pol ι (R. Woodgate, A. Tissier, J. P. McDonald, and E. G. Frank, personal communication; Table 1). Additionally, Luis Blanco and his colleagues (M. García-Díaz, O. Domínguez, L. A. López-Fernández, T. Laín de Lera, M. L. Saníger, J. F. Ruiz, M. Párraga, M. J. García-Ortiz, T. Kirchhoff, and J. del Mazo, personal communication, and ref. 33) have discovered two mammalian DNA polymerases called pol λ and pol μ, encoded by the *POLL* and *POLM* genes, respectively (Table 1). Mouse Pol λ shares 34% amino acid identity with DNA polymerase β but has an additional 200-amino acid region that includes a BRCT domain (M. García-Díaz, O. Domínguez, L. A. López-Fernández, T. Laín de Lera, M. L. Saníger, J. F. Ruiz, M. Párraga, M. J. García-Ortiz, T. Kirchhoff, and J. del Mazo, personal communication). This domain in other proteins (including some that are involved in cell-cycle checkpoint control and DNA repair) has been implicated in protein–protein interactions. Expression of mouse pol λ occurs predominantly in pachytene spermatocytes in the testis (M. García-Díaz, O. Domínguez, L. A. López-Fernández, T. Laín de Lera, M. L. Saníger, J. F. Ruiz, M. Párraga, M. J. García-Ortiz, T. Kirchhoff, and J. del Mazo, unpublished work).

Human DNA pol μ is 41% identical to deoxynucleotidyl terminal transferase, a well known enzyme that is not a true DNA polymerase, because, although it can add nucleotides at the end of a DNA template, it cannot extend a DNA primer in a DNA template-instructed manner. Human pol μ and deoxynucleotidyl terminal transferase are similar in size, and pol μ has deoxynucleotidyl terminal transferase activity (33). However, in contrast to deoxynucleotidyl terminal transferase, the activity of pol μ is strongly enhanced in the presence of a primer-template (33). The *POLM* gene is predominantly expressed in peripheral lymphoid tissues such as lymph nodes, the spleen, and the thymus, suggesting the exciting possibility that pol μ may be involved in somatic hypermutation, which facilitates maturation and expansion of the antibody repertoire.

Finally, the *D. melanogaster MUS308* gene, involved in the repair of DNA crosslinks (34), has been shown to encode a DNA polymerase (35). A human homolog of the *MUS308* gene called *POLQ* has been cloned recently (ref. 36; Table 1), and the polypeptide encoded by this gene has been shown to have DNA polymerase activity called pol θ (Table 1; F. S. Sharief and W. C. Copeland, unpublished observations).

What about the biological function of these most recent DNA polymerases? Preliminary studies indicate that like a number of the other DNA polymerases addressed above, pol μ has weak fidelity (33). Its ability (and that of pol λ and pol θ) to support TLS of various of types of base damage and structurally altered primer templates remains to be determined. As just mentioned, pol μ (and/or possibly one or more of the other DNA polymerases discussed herein) may play a role in somatic hypermutation, and pol λ may play a role in spermatogenesis. Interestingly, like the human *POLL* gene (M. García-Díaz, O. Domínguez, L. A. López-Fernández, T. Laín de Lera, M. L. Saníger, J. F. Ruiz, M. Párraga, M. J. García-Ortiz, T. Kirchhoff, and J. del Mazo, unpublished work), the mouse *Dinb1* and *Poli* genes are highly expressed in the testis in a cell lineage-specific manner. (ref. 15 and V.L.G., J. A. Richardson, and E.C.F., unpublished observations).

As is often the case in rapidly emerging fields, there is the potential for confusing the nomenclature of these polymerases and the genes that encode them. Human DinB1 protein has been designated as pol θ (22). However, as already indicated, pol θ is the previously published name of the product of the *POLQ* gene (35). Additionally, the products of yeast and human genes called *REV1* have all of the hallmarks of the DNA polymerases involved in TLS yet are named deoxycytidyl transferases (37, 38). Table 1 presents a modified nomenclature for the human DNA polymerases together with existing aliases in the literature, consistent with the long-established protocol of naming eukaryotic DNA polymerases consecutively according to the Greek alphabet (39, 40).

1. Walker, G. C. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10348–10350.
2. Woodgate, R. (1999) *Genes Dev.* **13**, 2191–2195.
3. Friedberg, E. C. & Gerlach, V. L. (1999) *Cell* **98**, 413–416.
4. Bridges, B. A. (1999) *Curr. Biol.* **9**, R475–R477.
5. Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 12224–12226.
6. Cleaver, J. E. (1999) *Science* **285**, 212–213.
7. Radman, M. (1999) *Nature* (*London*) **401**, 866–868.
8. Wood, R. D. (1999) *Nature* (*London*) **399**, 639–640.
9. Hübscher, U., Nasheuer, H.-P. & Syväoja, J. E. (2000) *Trends Biol. Chem.* **25**, 143–147.
10. Sander, M. & Samson, L. D. (2000) *Trends Cell Biol.* **10**, 159–162.
11. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J. & Marians, K. J. (2000) *Nature* (*London*) **404**, 37–41.
12. Bayton, K. & Fuchs, R. P. P. (2000) *Trends Biol. Chem.* **25**, 74–79.
13. Friedberg, E. C., Walker, G. C. & Siede, W. (1995) *DNA Repair and Mutagenesis* (Am. Soc. Microbiol., Washington, DC).
14. Gerlach, V. L., Aravind, L., Gotway, G., Schultz, R. A., Koonin, E. V. & Friedberg, E. C. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11922–11927.
15. McDonald, J. P., Rapić-Otrin, V., Epstein, J. A., Broughton, B. C., Wang, X., Lehmann, A. R., Wolgemuth, D. J. & Woodgate, R. (1999) *Genomics* **60**, 20–30.
16. Reuven, N. B., Tomer, G. & Livneh, Z. (1998) *Mol. Cell* **2**, 191–199.
17. Tang, M., Shen, X., Frank, E. G., O'Donnell, M., Woodgate, R. & Goodman, M. F. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 8919–8924.
18. Maor-Shoshani, A., Reuven, N. B., Tomer, G. & Livneh, Z. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 565–570.
19. Tang, M., Phan, P., Shen, X., Taylor, J.-S., O'Donnell, M., Woodgate, R. & Goodman, M. F. (2000) *Nature* (*London*) **404**, 1014–1018.
20. Wagner, J., Gruz, P., Kim, S.-R., Yamada, M., Yamamoto, Y., Matsui, K., Fuchs, R. P. P. & Nohmi, T. (1999) *Mol. Cell* **4**, 281–287.
21. Ogi, T., Kato, T., Jr., Kato, T. & Ohmori, H. (1999) *Genes Cells* **4**, 607–618.
22. Johnson, R. E., Prakash, S. & Prakash, L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3838–3843.
23. Matsutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S. & Hanaoka, F. (1999) *EMBO J.* **12**, 3491–3501.
24. Matsutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K. & Hanaoka, F. (1999) *Nature* (*London*) **399**, 700–704.
25. Johnson, R. E., Prakash, S. & Prakash, L. (1999) *Science* **283**, 1001–1004.
26. Johnson, R. E., Prakash, S. & Prakash, L. (1999) *J. Biol. Chem.* **274**, 15975–15977.
27. Johnson, R. E., Kondratick, C. M., Prakash, S. & Prakash, L. (1999) *Science* **285**, 263–265.
28. Yuan, F., Zhang, Y., Rajpal, D. K., Wu, X., Guo, D., Wang, M., Taylor, J.-S. & Wang, Z. J. (2000) *J. Biol. Chem.* **275**, 8233–8239.
29. Washington, M. T., Johnson, R. E., Prakash, S. & Prakash, L. (1999) *J. Biol. Chem.* **274**, 36835–36838.
30. Washington, M. T., Johnson, R. E., Prakash, S. & Prakash, L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3094–3099.
31. Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. (2000) *J. Biol. Chem.* **275**, 7447–7450.
32. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F. & Kunkel, T. A. (2000) *Nature* (*London*) **404**, 1011–1013.
33. Domínguez, O., Ruiz, J. F., de Lera, T. L., García-Díaz, M., González, M. A., Kirchkoff, T., Martínez, A. C., Bernad, A. & Blanco, L. (2000) *EMBO J.* **19**, 1731–1742.
34. Harris, P. V., Mazina, O. M., Leonhardt, E. A., Case, R. B., Boyd, J. B. & Burtis, K. C. (1996) *Mol. Cell. Biol.* **16**, 5764–5771.
35. Oshige, M., Aoyagi, N., Harris, P. V., Burtis, K. C. & Sakaguchi, K. (1999) *Mutat. Res.* **433**, 183–192.
36. Sharief, F. S., Vojta, P. J., Ropp, P. A. & Copeland, W. C. (1999) *Genomics* **59**, 90–96.
37. Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. (1996) *Nature* (*London*) **382**, 729–731.
38. Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F. & Wang, Z. (1999) *Nucleic Acids Res.* **27**, 4468–4475.
39. Weissbach, A., Baltimore, D., Bollum, F., Gallo, R. & Korn, D. (1975) *Science* **190**, 4014.
40. Burgers, P. M., Bambara, R. A., Campbell, J. L., Chang, L. M., Downey, K. M., Hübscher, U., Lee, M. Y., Linn, S. M., So, A. G. & Spadari, S. (1990) *Eur. J. Biochem.* **191**, 617–618.
41. Braithwaite, D. K. & Ito, J. (1993) *Nucleic Acids Res.* **21**, 787–802.
42. Aravind, L. & Koonin, E. V. (1999) *Nucleic Acids Res.* **27**, 1609–1618.