

Bridging the gap: A family of novel DNA polymerases that replicate faulty DNA

Robert E. Johnson, M. Todd Washington, Satya Prakash, and Louise Prakash*

Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, TX 77555-1061

DNA replication can be likened to a car travelling down a long highway. Normally the road is paved smooth, but, as often occurs during travel, the road sometimes contains bumps and roadblocks. Such is the case for the replication machinery as it travels down DNA. While DNA damaged by natural and manmade agents normally is repaired by a multitude of repair pathways, some lesions inevitably escape repair and, as a result, are encountered by the DNA replication machinery. Many of the lesions cannot be bypassed by the replicative DNA polymerases and will lead to cell death if not overcome. Just how cells handle these road blocks has been a longstanding problem.

Upon encountering a lesion, the replicative polymerase may dissociate from DNA, leaving a gap in the newly synthesized strand. Such a gap could be filled in by a recombinational mechanism (1) or by a "copy choice" type of DNA synthesis (2). Because both of these mechanisms use information from the undamaged sister duplex to fill in the gap, they are relatively error-free. However, replication of damaged DNA also can occur by synthesis across from the lesion in the template strand. In this case, a specialized replication complex inserts a random nucleotide across from the lesion and continues synthesis beyond the lesion. This process is usually mutagenic and is best exemplified by the *Escherichia coli* SOS system in which the complex of UmuC and UmuD' proteins interacts with DNA polymerase III (PolIII) holoenzyme to promote damage bypass. Until recently, the prevailing notion for UmuC-UmuD' action has been that it overrides the normally high fidelity of PolIII, enabling PolIII to insert a nucleotide across from the lesion and to replicate past the lesion (3).

UmuC belongs to a protein family that includes *E. coli* DinB, *Saccharomyces cerevisiae* Rev1 and Rad30, and human hRad30 proteins. In this issue of PNAS, Gerlach *et al.* (4) identify another member of this family in humans, which they refer to as DinB1. Recent studies of members of the UmuC protein family have heralded the emergence of a new paradigm of damage bypass in which, rather than modifying the activity of the replicative DNA polymerase, these pro-

teins themselves are DNA polymerases specific for bypassing lesions in a mutagenic or error-free manner (Table 1). The first hint of this new paradigm came from studies in *S. cerevisiae*, where genes belonging to the *RAD6* epistasis group function in error-free and error-prone damage bypass. The *REV1*, *REV3*, and *REV7* genes of this group are essential for mutagenic bypass, and *RAD5* and *RAD30* function in error-free bypass of UV-damaged DNA. Rev1 possesses an unusual deoxycytidyl transferase activity that specifically inserts a dCMP residue opposite a template abasic site (5). Rev1 also can insert a cytosine opposite template guanine, adenine, and uracil, but it does so with a much reduced efficiency (10–20%). The insertion of cytosine opposite the abasic site produces a terminus that can be extended by DNA polymerase ζ , comprised of the Rev3 and Rev7 subunits (6). Thus, mutagenic bypass of abasic sites in yeast could occur by the coordinated action of a transferase and an error-prone DNA polymerase (Fig. 1).

Although the deoxycytidyl transferase activity of Rev1 is template specific, Rev1 is clearly distinct from classical DNA polymerases in that it inserts only a cytosine residue. The first member of this family of proteins shown to be a bona fide, classical DNA polymerase was *S. cerevisiae* RAD30-encoded Pol η (7). Unlike Rev1, Pol η synthesizes DNA by incorporating all four nucleotides in a template-specific manner. Deletion of *RAD30* results in moderate sensitivity to UV light, and deletion of both *RAD5* and *RAD30* causes a synergistic increase in UV sensitivity and in UV mutagenesis, implicating these genes in alternate error-free bypass pathways (8–10). By contrast to error-prone synthesis by the Rev proteins, Pol η bypasses a thymine-thymine (T-T) dimer, a major UV photoproduct, in an error-free manner by inserting two A residues opposite the two Ts of the dimer (7) (Fig. 1).

The sun-sensitive cancer-prone syndrome xeroderma pigmentosum (XP) can arise from a defect in nucleotide excision repair (NER) or from a defect in the replication of UV-damaged DNA. Cells from the variant form of XP (XP-V) have no defect in NER, but they are unable to replicate UV-damaged DNA (11–13). Additionally, XP-V

cells are hypermutable with UV light, and they are less likely than normal cells to incorporate adenines opposite thymine photoproducts (14, 15). These observations, and the ability of the yeast enzyme to faithfully replicate dimer-containing DNA, prompted the proposal that mutations in human Pol η would cause XP-V (7).

The gene for the human *RAD30* counterpart, *hRAD30A*, recently was identified by two different groups and shown to harbor nonsense or frameshift mutations in cell lines derived from XP-V patients (16, 17). Because the majority of such mutations in *hRAD30A* would produce severe truncations of the protein and result in the loss of function, *hRAD30A* is not essential for viability or for growth. Like its yeast counterpart, *hRAD30A*-encoded Pol η bypasses a T-T dimer (17). In the absence of error-free replication by Pol η , error-prone translesion synthesis by Pol ζ would cause hypermutability and result in increased incidence of cancers that occurs in XP-V patients.

Another human *RAD30* homolog, *hRAD30B*, recently has been identified (18). The gene is located on chromosome 18q21.1, a region that frequently is deleted in many cancers. It remains to be seen whether mutations in this gene contribute to any of these cancers. The *hRAD30B* transcripts are highly expressed in testes and to a lesser extent in heart and pancreas. The human homolog of *E. coli* DinB, *hDINB1*, is located on chromosome 5q13.1. *hDINB1* is expressed in a variety of tissues, and its expression is also highest in testes (4). The high level expression in testes may reflect a specific role of these genes in some aspect of spermatogenesis or spermiogenesis.

In *E. coli*, mutagenic bypass of abasic sites recently has been reconstituted by using purified UmuC, UmuD', activated RecA, β -sliding clamp, γ -clamp loading complex, single-stranded DNA-binding protein

See companion article on page 11922 in issue 21 of volume 96.

*To whom reprint requests should be addressed at: Sealy Center for Molecular Science, University of Texas Medical Branch, 6.104 Medical Research Building, 11th and Mechanic Streets, Galveston, TX 77555-1061. E-mail: lprakash@scms.utmb.edu.

Table 1. Role of UmuC-related proteins

Subfamily	Cell type	Examples	Biochemical activity	Cellular function
UmuC	Prokaryotic	<i>E. coli</i> UmuC (PolV)	DNA polymerase	Error-prone damage bypass
DinB	Prokaryotic	<i>E. coli</i> DinB (PolIV)	DNA polymerase	Error-prone replication of undamaged DNA
	Eukaryotic	<i>S. pombe</i> SPCC5533.07c <i>C. elegans</i> F22B7.6 Human DinB1 Mouse DinB1		
Rad30A	Eukaryotic	<i>S. cerevisiae</i> Rad30 (Pol η) <i>S. pombe</i> SPBC16A3.11 Human Rad30A, XPV (Pol η)	DNA polymerase	Error-free bypass of UV damage
Rad30B	Eukaryotic	Human Rad30B Mouse Rad30B	Unknown	Unknown
Rev1	Eukaryotic	<i>S. cerevisiae</i> Rev1 <i>S. pombe</i> SPBC1347.01c <i>C. elegans</i> ZK675.2 Human CAB38231.1	Deoxycytidyl-transferase	Error-prone damage bypass

(SSB), and PolIII. In one study, even in the presence of β , γ -complex, SSB, and RecA, synthesis by PolIII was almost completely blocked one base before the lesion, and lesion bypass required the addition of the UmuD'₂C complex (19). In another study, PolIII holoenzyme (PolIII plus β , γ -complex), in the presence of RecA and SSB, could bypass the lesion, but it did so by skipping over the lesion, generating a 1-base deletion (20). Addition of UmuD' and UmuC to this reaction mix stimulated lesion bypass and also caused a dramatic shift in mutagenic specificity such that insertion of dAMP opposite the lesion was now favored, in agreement with the *in vivo* mutagenic specificity of abasic sites (20). From these observations, it was inferred that the UmuD'₂C complex prevents misalignment of the lesion and favors the formation of a

stable multiprotein complex at the template-primer junction.

DNA polymerase activity now has been identified in the UmuD'₂C complex as well (21). Although the UmuD'₂C complex alone shows little ability to insert a nucleotide across from the abasic site and to extend from it, lesion bypass occurs in the presence of β , γ -complex, single-stranded DNA-binding protein, and RecA. Because the UmuD'₂C polymerase activity, designated PolV, is highly distributive, PolIII holoenzyme must take over from UmuD'₂C once bypass has occurred. Thus, replication of damaged DNA in *E. coli* would involve the action of the UmuD'₂C complex for replication across and beyond the lesion (Fig. 1) and of PolIII holoenzyme for continued synthesis.

A UmuC homolog in *E. coli*, *dinB*, initially was identified in a screen for genes

whose expression is stimulated by DNA damage (22). By contrast to the UmuD'₂C-dependent pathway that is essential for mutagenesis by UV and other DNA-damaging agents, *dinB* is involved in untargeted mutagenesis of phage λ DNA when undamaged phage infects UV-irradiated *E. coli* cells. Additionally, overexpression of *dinB* results in a dramatic increase in spontaneous mutagenesis of F' lac plasmids, primarily causing -1 deletions in homopolymeric runs (22). Thus, *dinB* can generate mutations in DNA that has not been exposed to a DNA-damaging agent. Recently, DinB also has been shown to contain a polymerase activity, designated PolIV, which can extend DNA from a misaligned primer-template (23). When a primer-template containing a G-G mismatch within a small repetitive sequence is used, DinB yields

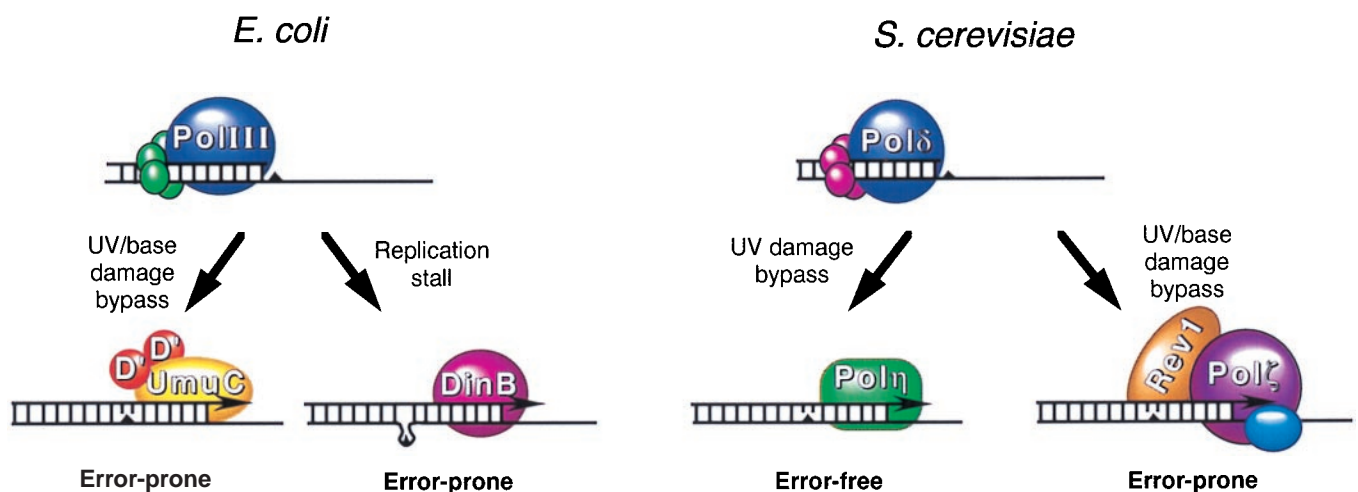


Fig. 1. Modes of bypass replication in *E. coli* and *S. cerevisiae*. As DNA replication proceeds, the replicative DNA polymerases, PolIII in *E. coli* or Pol δ in *S. cerevisiae*, encounter lesions or stall sites (triangles) in DNA. In *E. coli*, the DNA polymerase activity of the UmuD'₂C complex carries out limited DNA synthesis across and past the lesion in an error-prone manner. At replicative pause sites, such as at a misaligned template-primer junction, the DinB protein could carry out limited synthesis, resulting in -1 frameshifts. In *S. cerevisiae*, the RAD30-encoded Pol η performs replicative bypass of thymine dimers in an error-free manner by the insertion of two adenines across from the dimer. Alternatively, the Rev1 protein together with the REV3/REV7-encoded Pol ζ carries out replicative bypass of UV lesions and other base damages in an error-prone manner. In humans, defects in Pol η in XP-V patients result in the loss of this error-free component of UV damage bypass. Elevated mutagenesis arising from increased bypass by Pol ζ would be the underlying cause of high cancer incidence in XP-V patients.

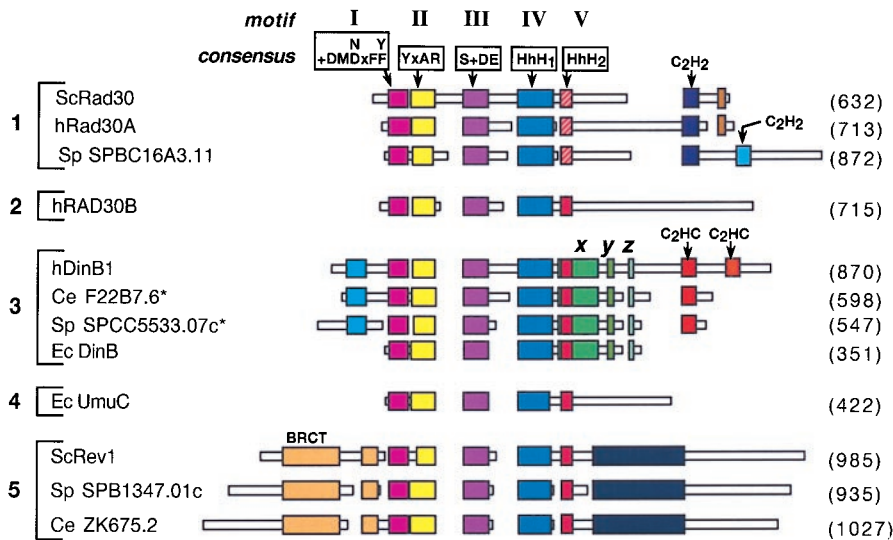


Fig. 2. Schematic alignment of DinB/UmuC/Rad30 family proteins. Boxes represent amino acid sequences of the proteins. Regions of homology are indicated by colored larger boxes, whereas unique sequences are shown as narrow white boxes. Areas without boxes indicate gaps introduced for optimal alignment. Motifs I-III are shown with their respective consensus sequences where x indicates any amino acid and + indicates a hydrophobic residue (I, L, or V). Motifs IV and V represent the helix-hairpin-helix (HhH) domains. The HhH₂ sequence in the hRad30A family, indicated in hatched red, differs from the analogous sequence present in the other proteins, indicated in solid red. The DinB-specific sequences are indicated by x, y and z. Zinc binding motifs are indicated as C₂H₂ or C₂HC. Numbers in parentheses indicate protein length in amino acids. Sc, *S. cerevisiae*; Sp, *S. pombe*; h, human; Ec, *E. coli*; Ce, *C. elegans*; BRCT, BRCA1 C-terminal domain. The asterisks on Ce F22B7.6 and Sp SPCC5533.07C indicate that these protein sequences were derived from the alternate predicted mRNA splice sites in the 3' regions of these genes that would produce a longer protein. Alignments were generated by using the MACAW and CLUSTAL W programs.

products that are 1 or 2 nt shorter than the ones generated by the Klenow fragment. DinB may promote replication through pause sites in repetitive DNA sequences via template misalignment (Fig. 1).

Phylogenetic comparisons have indicated the existence of four subfamilies within this family of DNA polymerases. These subfamilies have been categorized as UmuC-like, DinB-like, Rad30-like, and Rev1-like (4). Although UmuC-like proteins have been identified only in Gram-negative and Gram-positive bacteria, DinB-like proteins have been found in both prokaryotic and eukaryotic organisms. The Rad30 and Rev1 subfamily of proteins has been found only in eukaryotic organisms. Alignment of some of these proteins from such diverse organisms as *E. coli*, *S. cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, and humans (Fig. 2) reveals the presence of the several highly conserved motifs, designated I-V. These motifs are likely involved in DNA binding and nucleotidyl transferase activities. In fact, mutations of acidic residues in motifs I and III in several of these proteins have been shown to inactivate the DNA polymerase activity (10, 21, 23). These alignments also show the presence of sub-

family-specific conserved motifs that lie outside motifs I-V. For instance, in the hRad30A family, there is a highly conserved C₂H₂ zinc finger motif in the C terminus, and although the C termini of DinB1 subfamily of eukaryotic proteins also contain a zinc-binding motif, it is distinct from that found in hRad30A protein family (Fig. 2). The Rev1 subfamily contains highly conserved N-terminal and C-terminal extensions that are not found in other subfamilies. The unique C-terminal or N-terminal extensions present in these protein subfamilies could function in specific protein-protein interactions.

Although phylogenetic analyses place the newly identified hRad30B protein in the Rad30-like subfamily (18), a comparison of sequences indicates that the C terminus of this protein is distinct from any other protein in the various subfamilies. In addition, hRad30B contains a divergent GFDE sequence in motif III, as opposed to the highly conserved S(I/L/V)DE sequence found in other members of this family. Because hRad30B has been found only in humans and mice, (the entire genomes of *S. cerevisiae* and *C. elegans* have been sequenced and no or-

tholog has been identified), hRad30B may represent a new subfamily of proteins distinct from the Rad30A, DinB, UmuC, and Rev1 subfamilies.

The presence of DNA polymerase activities in the above-mentioned proteins has revealed a novel mechanism of damage bypass, conserved among prokaryotes and eukaryotes. These studies now raise a number of other questions that need to be addressed. For example: How are these alternative DNA polymerases coupled to the replicative machinery? Do these enzymes function in isolation, or are they part of multisubunit protein complexes? What is the range of DNA lesions that yeast and human Pol η can bypass? Is Pol η specific for the bypass of a T-T dimer, or can it bypass other UV lesions such as a T-C, C-T, and C-C dimer, and is Pol η an "A" rule polymerase that bypasses some lesions correctly and others incorrectly? And finally, what are the roles of human Rad30B and DinB1 proteins? Do they function in error-free or in mutagenic damage bypass, and are they specific for different types of DNA lesions? It is hoped that answers to these and related questions will be forthcoming soon.

1. Rupp, W. D., Wilde, C. E. I., Reno, D. L. & Howard-Flanders, P. (1971) *J. Mol. Biol.* **61**, 25-44.
2. Higgins, N. P., Kato, K. & Strauss, B. (1976) *J. Mol. Biol.* **101**, 417-425.
3. Friedberg, E. C., Walker, G. C. & Siede, W. (1995) *DNA Repair and Mutagenesis* (Am. Soc. Microbiol. Press, Washington, DC).
4. 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.
5. Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. (1996) *Nature (London)* **382**, 729-731.
6. Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. (1996) *Science* **272**, 1646-1649.
7. Johnson, R. E., Prakash, S. & Prakash, L. (1999) *Science* **283**, 1001-1004.
8. McDonald, J. P., Levine, A. S. & Woodgate, R. (1997)

- Genetics* **147**, 1557-1568.
9. Roush, A. A., Suarez, M., Friedberg, E. C., Radman, M. & Siede, W. (1998) *Mol. Gen. Genet.* **257**, 686-692.
10. Johnson, R. E., Prakash, S. & Prakash, L. (1999) *J. Biol. Chem.* **274**, 15975-15977.
11. Lehmann, A. R., Kirk-Bell, S., Arlett, C. F., Paterson, M. C., Lohman, P. H. M., de Weerd-Kastelein, E. A. & Bootsma, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 219-223.
12. Cleaver, J. E., Arutyunyan, R. M., Sarkisian, T., Kaufmann, W. K., Greene, A. E. & Coriell, L. (1980) *Carcinogenesis* **1**, 647-655.
13. Cordeiro-Stone, M., Zaritskaya, L. S., Price, L. K. & Kaufmann, W. K. (1997) *J. Biol. Chem.* **272**, 13945-13954.
14. Wang, Y.-C., Maher, V. M., Mitchell, D. L. & McCormick, J. J. (1993) *Mol. Cell. Biol.* **13**, 4276-4283.
15. Waters, H. L., Seetharam, S., Seidman, M. M. & Kraemer, K. H. (1993) *J. Invest. Dermatol.* **101**, 744-748.

16. Johnson, R. E., Kondratik, C. M., Prakash, S. & Prakash, L. (1999) *Science* **285**, 263-265.
17. Masutani, 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.
18. McDonald, J. P., Rapic-Otrin, V., Epstein, J. A., Broughton, B. C., Wang, X., Lehmann, A. R., Wolgemuth, D. J. & Woodgate, R. (1999) *Genomics* **60**, 20-30.
19. Tang, M., Bruck, I., Eritja, R., Turner, J., Frank, E. G., Woodgate, R., O'Donnell, M. & Goodman, M. F. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 9755-9760.
20. Reuven, N. B., Tomer, G. & Livneh, Z. (1998) *Mol. Cell* **2**, 191-199.
21. Tang, M., Shen, X., Frank, E. G., O'Donnell, M., Woodgate, R. & Goodman, M. F. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 8919-8924.
22. Smith, B. T. & Walker, G. C. (1998) *Genetics* **148**, 1599-1610.
23. Wagner, J., Gruz, P., Kim, S.-R., Yamada, M., Matsui, K., Fuchs, R. P. P. & Nohmi, T. (1999) *Mol. Cell* **4**, 281-286.