Evolving Views of DNA Replication (In)Fidelity

T.A. KUNKEL

Laboratory of Molecular Genetics and Laboratory of Structural Biology, National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, North Carolina 27709

Correspondence: kunkel@niehs.nih.gov

“It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material” (Watson and Crick 1953).

In the years since this remarkable understatement, we have come to realize the enormous complexity of the cellular machinery devoted to replicating DNA with the accuracy needed to maintain genetic information over many generations, balanced by the emergence of mutations on which selection can act. This complexity is partly based on the need to remove or tolerate cytotoxic and mutagenic lesions in DNA generated by environmental stress. Considered here is the fidelity with which undamaged and damaged DNA is replicated by the many DNA polymerases now known to exist. Some of these seriously violate Watson–Crick base-pairing rules such that, depending on the polymerase, the composition and location of the error, and the ability to correct errors (or not), DNA synthesis error rates can vary by more than a millionfold. This offers the potential to modulate rates of point mutations over a wide range, with consequences that can be either deleterious or beneficial.

In organisms from viruses to man, the fidelity with which genetic information is replicated depends on the ability of polymerases to select correct nucleotides—rather than incorrect and/or damaged nucleotides—for incorporation without adding or deleting nucleotides. Polymerase selectivity is the prime determinant of fidelity both at the replication fork and during synthesis to repair DNA damage generated by endogenous cellular metabolism or exposure to the environment (Friedberg et al. 2006). In many organisms, fidelity can be increased by exonucleolytic proofreading of mismatches during replication and by DNA mismatch repair (MMR) (for review, see Kunkel and Erie 2005; Iyer et al. 2006; Hsieh and Yamane 2008). Certain proteins involved in MMR also can also signal for DNA-damage responses, prevent homologous recombination, promote meiotic recombination, modulate somatic hypermutation of immunoglobulin genes, or even stabilize certain misaligned repetitive DNA sequences. When DNA damage is not removed before replication, helix-distorting lesions can impede replication fork progression. In such circumstances, cell survival can be enhanced by specialized DNA transactions, some of which can be mutagenic via translesion DNA synthesis (Jansen et al. 2007; Yang and Woodgate 2007; Chang and Cimprich 2009). Considered here is the amazing diversity of evolutionarily conserved DNA polymerases involved in these transactions, many of which have been discovered relatively recently. Emphasis is on their fidelity and on the contributions of proofreading and MMR to replication fidelity, which can vary over a much wider range than was appreciated even a decade ago.

MUTIPLE POLYMERASES WITH MULTIPLE OVERLAPPING FUNCTIONS

DNA polymerases were first discovered using assays for polymerization activity (Bessman et al. 1956). This approach revealed that bacteria and eukaryotes harbor multiple polymerases (Kornberg and Baker 1992). However, just how many only came to light more recently when sequence alignments and recombinant DNA technology were used to find low-activity low-abundance polymerases. Sequence alignments now permit classification of DNA polymerases into several different families, with most organisms encoding more than one (Shcherbakova et al. 2003a; Bebenek and Kunkel 2004; Loeb and Monnat 2008). For example, Escherichia coli encodes five polymerases (Friedberg et al. 2005), one each from families A, B, and C and two from different subfamilies of family Y, each with important but somewhat different functions. The human genome encodes even more (Table 1) from families A (three polymerases), B (four polymerases), X (four polymerases), Y (four polymerases), and reverse transcriptase (RT) (telomerase). Because polymerases can have multiple functions (Table 1) and can sometimes compensate one for another, it is a continuing challenge to understand exactly where and when each polymerase operates in vivo.

Despite differences in primary sequence, DNA polymerases in different families share a common general structure for the polymerase domain (Olins et al. 1985), which is composed of fingers, thumb, and palm subdomains (colored blue, green, and red, respectively, in Fig. 1A). The palm contains three highly conserved carboxy-
**Table 1.** Human DNA-template-dependent DNA Polymerases

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>Family</th>
<th>Mass (kDa)</th>
<th>Gene (alias)</th>
<th>Associated activities</th>
<th>Proposed functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>B</td>
<td>165</td>
<td>POLA</td>
<td>RNA primase</td>
<td>nuclear genome replication, S-phase checkpoint</td>
</tr>
<tr>
<td>β</td>
<td>X</td>
<td>39</td>
<td>POLB</td>
<td>dRP lyase</td>
<td>BER</td>
</tr>
<tr>
<td>γ</td>
<td>A</td>
<td>140</td>
<td>POLG</td>
<td>3′→5′ exonuclease</td>
<td>mitochondrial genome replication and BER</td>
</tr>
<tr>
<td>δ</td>
<td>B</td>
<td>125</td>
<td>POLD1</td>
<td>3′→5′ exonuclease</td>
<td>nuclear genome replication NER, BER, MMR, DSB repair</td>
</tr>
<tr>
<td>ε</td>
<td>B</td>
<td>255</td>
<td>POLE</td>
<td>3′→5′ exonuclease</td>
<td>nuclear genome replication NER, BER, MMR, DSB repair</td>
</tr>
<tr>
<td>ζ</td>
<td>B</td>
<td>353</td>
<td>POLZ (REV3)</td>
<td>–</td>
<td>TLS, DSB repair, ICL repair, SHM</td>
</tr>
<tr>
<td>η</td>
<td>Y</td>
<td>78</td>
<td>POLH (RAD30, RAD30A, XPV)</td>
<td>–</td>
<td>TLS, SHM, BER?</td>
</tr>
<tr>
<td>θ</td>
<td>A</td>
<td>198</td>
<td>POLQ</td>
<td>helicase motifs</td>
<td>ICL repair? TLS, SHM?</td>
</tr>
<tr>
<td>ι</td>
<td>Y</td>
<td>80</td>
<td>POL1 (RAD30B)</td>
<td>dRP lyase</td>
<td>TLS, BER SHM?</td>
</tr>
<tr>
<td>κ</td>
<td>Y</td>
<td>76</td>
<td>POLK (DINB1)</td>
<td>–</td>
<td>TLS</td>
</tr>
<tr>
<td>λ</td>
<td>X</td>
<td>66</td>
<td>POLL</td>
<td>dRP lyase</td>
<td>V(D)J recombination</td>
</tr>
<tr>
<td>μ</td>
<td>X</td>
<td>55</td>
<td>POLM</td>
<td>terminal transferase</td>
<td>V(D)J recombination</td>
</tr>
<tr>
<td>ν</td>
<td>A</td>
<td>100</td>
<td>POLN</td>
<td>–</td>
<td>TLS?</td>
</tr>
<tr>
<td>σ</td>
<td>X</td>
<td>60</td>
<td>POLS (TRF4-1)</td>
<td>3′→5′ exonuclease</td>
<td>sister chromatid cohesion</td>
</tr>
<tr>
<td>TdT</td>
<td>X</td>
<td>58</td>
<td>TdT</td>
<td>–</td>
<td>V(D)J recombination</td>
</tr>
<tr>
<td>REV1</td>
<td>Y</td>
<td>138</td>
<td>REV1</td>
<td>dCTP incorporation</td>
<td>TLS</td>
</tr>
</tbody>
</table>


**Figure 1.** X-ray crystal structures of DNA polymerases. (A) Shown is the structure of a representative replicative DNA polymerase from bacteriophage RB69 (family B). Polymerase domains share three common subdomains: designated fingers (blue), palm (red), and thumb (green). Other domains for specialized functions are shown in purple and yellow. (B) The active site of human DNA polymerase β. The surface of Arg-283 is highlighted in pink to emphasize the importance to fidelity of polymerase interactions with the DNA minor groove. (C) The more open and solvent-accessible active site of low-fidelity *Sulfolobus sulfataricus* Dpo4. See text for further descriptions. (A, Prepared by Miguel Garcia-Diaz, using the structure in Franklin et al. [2001]; B and C, reprinted, with permission, from Kunkel et al. 2003.)
lates that bind two divalent metal ions required for catalysis via an in-line nucleophilic attack of the 3′-OH on the α-phosphate of the incoming dNTP. This mechanism is thought to be common to all DNA polymerases (Steitz 1993), yet it appears to have resulted from convergent evolution, because some polymerase families have the active site carboxylates in a “right-handed” configuration and others (families X and C) have it in a “left-handed” configuration (see, e.g., Wang et al. 2008 and references therein).

The polymerase domains are usually attached to other domains needed for the variable functions of these proteins. For example, polymerases that perform the bulk of genome replication often have a domain harboring 3′ exonuclease activity that proofreads replication errors (Fig. 1A). Nonetheless, most DNA polymerases lack an intrinsic 3′ exonuclease activity (Table 1), which is interesting given the importance of proofreading to genome stability (see below). Other specialized domains include a “little finger” domain (Yang and Woodgate 2007) unique to family-Y members involved in translesion DNA synthesis, and an 8-kDa domain unique to family-X polymerases (Moon et al. 2007) that assists in filling small gaps during DNA repair and that, in polymerases 8 and λ, harbors a dRP (deoxyribose phosphate) lyase activity needed for repair (Table 1). Still other domains include the BRCT (BRCA1 carboxy terminal) domains of family-X polymerases involved in nonhomologous end-joining of double-stranded DNA breaks and amino-or carboxy-terminal regions of polymerase catalytic subunits that are involved in cellular responses to DNA damage, including via partnerships with other proteins. In fact, DNA polymerases typically operate in DNA transactions that require coordinated interactions with many other proteins (e.g., noncatalytic accessory subunits, processivity clamps, and single-stranded DNA-binding proteins), whose properties and functions are subjects of continuing interest (see, e.g., Shcherbakova et al. 2003a; Bebenek and Kunkel 2004; Friedberg et al. 2005; Jansen et al. 2007; Loeb and Monnat 2008; Burgers 2009; Chang and Cimprich 2009 and references therein).

The polymerase domains are usually attached to other domains needed for the variable functions of these proteins. For example, polymerases that perform the bulk of genome replication often have a domain harboring 3′ exonuclease activity that proofreads replication errors (Fig. 1A). Nonetheless, most DNA polymerases lack an intrinsic 3′ exonuclease activity (Table 1), which is interesting given the importance of proofreading to genome stability (see below). Other specialized domains include a “little finger” domain (Yang and Woodgate 2007) unique to family-Y members involved in translesion DNA synthesis, and an 8-kDa domain unique to family-X polymerases (Moon et al. 2007) that assists in filling small gaps during DNA repair and that, in polymerases 8 and λ, harbors a dRP (deoxyribose phosphate) lyase activity needed for repair (Table 1). Still other domains include the BRCT (BRCA1 carboxy terminal) domains of family-X polymerases involved in nonhomologous end-joining of double-stranded DNA breaks and amino-or carboxy-terminal regions of polymerase catalytic subunits that are involved in cellular responses to DNA damage, including via partnerships with other proteins. In fact, DNA polymerases typically operate in DNA transactions that require coordinated interactions with many other proteins (e.g., noncatalytic accessory subunits, processivity clamps, and single-stranded DNA-binding proteins), whose properties and functions are subjects of continuing interest (see, e.g., Shcherbakova et al. 2003a; Bebenek and Kunkel 2004; Friedberg et al. 2005; Jansen et al. 2007; Loeb and Monnat 2008; Burgers 2009; Chang and Cimprich 2009 and references therein).

**Figure 2.** Polymerase error rates and the contributions of each fidelity process to mutation rate. The image illustrates the wide ranges over which polymerase nucleotide selectivity, exonucleolytic proofreading, and mismatch repair contribute to spontaneous mutation rates of organisms. Also depicted are the average rates at which purified eukaryotic DNA polymerases generate single-base substitution and single-base deletion errors when performing gap-filling DNA synthesis in vitro. See text for further descriptions. For details on the source and composition of the polymerases used and on their error specificity, see McCulloch and Kunkel (2008 and references therein).
they decrease the entropy of the system. This magnifies the contribution of enthalpy to the free-energy difference (Petruka and Goodman 1995), thereby increasing nucleotide selectivity. Another idea supported by substantial evidence (for review, see Kunkel and Bebenek 2000; Kool 2002; Beard and Wilson 2003; Kim et al. 2005) is that high nucleotide selectivity partly results from the shape complementarity in the nascent base-pair-binding pocket. The four canonical Watson–Crick base pairs are nearly identical in size and shape. Structural studies reveal that correct base pairs fit within the nascent base-pair-binding pocket without steric clashes. Particularly important to fidelity are amino acid side chains (e.g., Fig. 1B, Arg-283 [purple] in polymerase β) that interact with the O2 atom of pyrimidines and the N3 atom of purines, which are isosteric in the four correct Watson–Crick base pairs. This is illustrated in Figure 1B, which shows the active site of DNA polymerase β, a relatively accurate repair enzyme, with a correct base pair poised for catalysis. The correct pair fits snugly, whereas mismatches with different and variable geometries are predicted to have steric clashes that would reduce incorrect dNTP-binding affinity, affect subsequent conformational changes needed to set up the proper geometry for catalysis, and/or reduce the rate of phosphodiester bond formation.

### Insertion–Deletion Errors (Indels)

DNA polymerases also insert and delete nucleotides during DNA synthesis. These errors result from strand misalignments that generate unpaired bases in the primer strand, leading to additions, or in the template strand, leading to deletions. Ideas to account for how these misalignments initiate and are stabilized for continued synthesis include classical DNA strand slippage, misinsertion followed by primer relocation, and misalignment of a nucleotide at the polymerase active site. Biochemical and structural support exists for all three models (for review, see Bebenek and Kunkel 2000; Garcia-Diaz and Kunkel 2006). Replicative DNA polymerases generate single-base deletions at rates that are similar to those for single-base substitutions (Fig. 2). Single-base deletion error rates are usually higher than single-base addition error rates or rates for indels involving large numbers of nucleotides, with possible explanations considered elsewhere (Bebenek and Kunkel 2000; Garcia-Diaz and Kunkel 2006). Importantly, the single-base substitution and deletion error rates in Figure 2 are average values, with wide variations observed depending on the type of mismatch and the sequence context in which the mismatch is located (Kunkel and Bebenek 2000). Prime examples of such variability among lower fidelity polymerases involved in DNA repair and translesion synthesis are considered below.

### Proofreading by Replicative DNA Polymerases

Average base substitution error rates of proofreading-proficient replicative DNA polymerases are typically $\geq 10^{-6}$. Their exonuclease-deficient derivatives are considerably less accurate, indicating that on average, proofreading improves replication fidelity by about ~10-fold to 100-fold (Fig. 2). The energetic cost of improving fidelity by more than this could be unacceptable due to excessive excision of correctly paired bases (Fersht et al. 1982). The biological importance of proofreading is illustrated by studies showing that when highly conserved residues near the active sites of *Saccharomyces cerevisiae* replicative polymerases are replaced with nonconservative amino acids, the mutant enzymes have decreased DNA synthesis fidelity in vitro (Longley et al. 2001; Scherbakova et al. 2003b; Fortune et al. 2005) and generate mutator phenotypes in vivo (Morrison and Sugino 1994). Moreover, mice with homologous replacements in polymerase δ have decreased genomic stability and accelerated tumorigenesis (Goldsbey et al. 2001).

The key to proofreading efficiency is the balance between polymerization and excision at a growing primer terminus (Fig. 3A). Under normal circumstances, correct incorporation allows subsequent incorporations to occur rapidly with little opportunity for proofreading (line 1). However, misinsertion generates a mismatched primer terminus that is more difficult to extend. This slows polymerization, allowing the primer terminus to fray and move single-stranded DNA into the exonuclease active site for excision of the error (line 2). On the basis of early work (for review, see Kornberg and Baker 1992) and on more recent studies, we now realize that there are several ways to influence this critical balance between polymerization and excision (Table 2). Proofreading can be inactivated by amino acid substitutions in the exonuclease active site, or exonuclease activity can be inhibited if the end product of excision, a dNMP, binds to the exonuclease active site. Proofreading can be reduced by amino acid substitutions in replicative polymerases that prevent movement of the frayed primer terminus to the exonuclease active site (so-called “switching mutants”) (see, e.g., Jin et al. 2005 and references therein) or by amino acid substitutions in the polymerase active site that promote mismatch extension (see, e.g., Nick McElhinny et al. 2008 and references therein). Proofreading can be suppressed by increasing the concentration of the next correct nucleotide to be incorporated after a misinsertion (dCTP for the examples in Fig. 3A), thereby promoting mismatch extension at the expense of excision (Ninio 1975). Finally, under some circumstances, mismatches escape proofreading by tricking the replicative polymerases. A well-known example involves 8-oxo-guanine, a common lesion resulting from oxidative stress. Replication of template 8-oxo-G can generate 8-oxoGdA mismatches whose geometry is similar to that of a correct base pair, such that the mismatch largely escapes proofreading, e.g., by replicative T7 DNA polymerase (see, e.g., Brieba et al. 2004). Another example with high biological relevance involves proofreading of insertion–deletion mismatches during replication of repetitive sequences (Fig. 3A, line 3). Proofreading does correct misaligned intermediates containing extra bases in one strand or the other near the primer terminus, as illustrated by the higher indel error rates of exonuclease-deficient polymerases δ, ε, and γ when compared to their proofreading-proficient
counterparts (Longley et al. 2001; Shcherbakova et al. 2003b; Fortune et al. 2005). However, the efficiency of indel proofreading decreases as the length of a repetitive sequence increases (e.g., Fig. 3B). This is because in a long repetitive sequence, the mismatch generated by strand slippage (i.e., the unpaired base) is likely to be located upstream of the polymerase active site, such that it does not strongly reduce the rate of polymerization (Fig. 3A, line 3). Such diminished proofreading, in conjunction with a higher rate of strand slippage by polymerases (e.g., left panel in Fig. 3B; see Garcia-Diaz and Kunkel 2006), contributes to the observation that long repetitive sequences are at risk for a high rate of replication slippage errors, as evidenced by the well-known “microsatellite
instability” phenotype of eukaryotic cells defective in DNA mismatch (see below). On the basis of the logic in Figure 3 and the parameters in Table 2, it is now very clear that just as for nucleotide selectivity, the contribution of proofreading to replication fidelity can vary over a wide range (Fig. 2), from almost none (8-oxoG•dA mismatches) to several hundredfold (e.g., for bacteriophage T7 replication; see, e.g., Donlin et al. 1991).

**“Extrinsic” Proofreading May Also Contribute to Genome Stability**

Interestingly, among many mammalian DNA polymerases, only those responsible for the bulk of chain elongation during replication (δ, ε, and γ) contain intrinsic 3′ exonucleolytic proofreading activity. Nonetheless, the exonuclease-deficient polymerases have very important roles in maintaining genome stability (Table 1). Are errors made by exonuclease-deficient polymerases subject to “extrinsic” proofreading by a separate exonuclease? The idea (Fig. 3A, line 4) is that, after making a mismatch, the polymerase would dissociate, allowing the exonuclease activity of another protein to excise the mismatch. Indeed, the major exonuclease activity of another protein to excise the mismatch, the polymerase would dissociate, allowing the exonuclease activity of another protein to excise the mismatch. On the basis of the logic in Figure 3 and the parameters in Table 2, it is now very clear that just as for nucleotide selectivity, the contribution of proofreading to replication fidelity can vary over a wide range (Fig. 2), from almost none (8-oxoG•dA mismatches) to several hundredfold (e.g., for bacteriophage T7 replication; see, e.g., Donlin et al. 1991).

**“Extrinsic” Proofreading May Also Contribute to Genome Stability**

Interestingly, among many mammalian DNA polymerases, only those responsible for the bulk of chain elongation during replication (δ, ε, and γ) contain intrinsic 3′ exonucleolytic proofreading activity. Nonetheless, the exonuclease-deficient polymerases have very important roles in maintaining genome stability (Table 1). Are errors made by exonuclease-deficient polymerases subject to “extrinsic” proofreading by a separate exonuclease? The idea (Fig. 3A, line 4) is that, after making a mismatch, the polymerase would dissociate, allowing the exonuclease activity of another protein to excise the mismatch. Indeed, the major exonuclease activity of another protein to excise the mismatch, the polymerase would dissociate, allowing the exonuclease activity of another protein to excise the mismatch. On the basis of the logic in Figure 3 and the parameters in Table 2, it is now very clear that just as for nucleotide selectivity, the contribution of proofreading to replication fidelity can vary over a wide range (Fig. 2), from almost none (8-oxoG•dA mismatches) to several hundredfold (e.g., for bacteriophage T7 replication; see, e.g., Donlin et al. 1991).

**REPLICATION ASYMMETRY AND FIDELITY**

The two strands of duplex DNA are oriented antiparallel to each other, and DNA polymerases copy DNA in only the 5′ to 3′ direction. Thus, replication of duplex DNA is intrinsically asymmetric. This asymmetry is illustrated by the simple model of a eukaryotic replication fork shown in Figure 4A, left. Recent evidence in budding yeast suggests that the leading strand is primarily replicated by polymerase ε (Pursell et al. 2007), whereas Okazaki fragments on the lagging strand are initiated by polymerase δ-primase and then primarily completed by polymerase δ (Nick McElhinny et al. 2008). These enzymes differ from each other in primary sequence, subunit composition, interactions with other proteins, and several biochemical properties, including processivity, proofreading capacity, fidelity, and error specificity. It is therefore possible that the fidelity of leading- and lagging-strand replication may differ, perhaps even more so under nonstandard replication conditions arising under stress (Fig. 4A, right), either environmental (DNA lesions) or genetic (mutations in key genes). Evidence for differences in leading- and lagging-strand replication fidelity have been reported in E. coli (Fijalkowska et al. 1998), where both strands are replicated by the same polymerase acting as a multisubunit dimer, DNA polymerase III holoenzyme.

**FIDELITY OF DNA REPAIR POLYMERASES**

Efficient and accurate replication requires clean substrates, such that many organisms devote great attention and energy to repairing DNA lesions that can result from endogenous metabolic processes and from exposure to physical and chemical agents in the external environment. Many different repair processes exist and can be distinguished by lesion specificity, the enzymes involved, and when they operate (for review, see Friedberg et al. 2006). For many of these repair pathways, e.g., BER, NER, NHEJ,
MMR, and ICL repair, excision of a lesion is followed by gap-filling DNA synthesis and ligation (Friedberg et al. 2006). The gap filling is conducted by polymerases that are highly accurate (e.g., polymerase δ for filling long gaps during NER and MMR) or moderately accurate (polymerase β for filling short gaps during BER). However, certain gap-filling transactions in cells may involve inaccurate DNA polymerases, e.g., κ in NER, ζ/θ in ICL repair, and µ in NHEJ, such that DNA synthesis errors occurring during repair may contribute to mutagenesis. Perhaps the best examples are for polymerases ζ and η, both of which are implicated in somatic hypermutation of immunoglobulin genes, a process that involves processing uracil in DNA generated by cytosine deamination catalyzed by activation-induced cytosine deaminase (Diaz and Lawrence 2005).

Fidelity of Translesion Synthesis Polymerases

Replication forks can stall upon encountering lesions that distort helix geometry. Among several possible solutions that allow complete genome replication, one is translesion synthesis (TLS) by DNA polymerases. Two general models have been put forth for TLS, one involving a single TLS polymerase for lesion bypass (Fig. 4B, left) and another involving two TLS polymerases, one for insertion opposite a lesion and another for extending aberrant primer termini (Fig. 4B, right). Several specialized TLS polymerases have been discovered in the past decade, and these are evolutionarily conserved (Ohmori et al. 2001; Prakash and Prakash 2002; Yang and Woodgate 2007). In mammals, they include family-B polymerase ζ, family-Y polymerases η, κ, and ι, and possibly family-A polymerases θ and v. These are the least accurate of polymerases (Fig. 2). They all lack proofreading activity and also have lower nucleotide selectivity than the major replicative polymerases, as indicated by their higher error rates for base substitutions and indels (Fig. 2). The extreme case is for polymerase ι, a conserved family-Y member that rarely generates certain mismatches (e.g., A•dC) but can preferentially misincorporate dG as compared to dA opposite template T (Fig. 2) (see, e.g., Bebenek et al. 2001). This rather amazing violation of Watson–Crick base-pairing dogma leads one to wonder what the true physiological substrates and functions of polymerase ι might be in vivo, whether in TLS (Dumstorf et al. 2006) or in yet to be discovered DNA transactions. Structural and biochemical studies suggest that the low fidelity of family-Y enzymes is partly due to relaxed geometric selectivity in the nascent base-pair-binding pocket, which is more open and solvent accessible than those of more accurate DNA polymerases. An example is shown in Figure 1C, which depicts the active site of a bacterial Y-family polymerase: Sso Dpo4. Indeed, much of the seminal work on family-Y polymerases has been performed using bacterial enzymes, which include two DNA-damage-induced E. coli DNA polymerases: IV and V (Fuchs et al. 2004). Another TLS polymerase is the B-family member ζ. When copying undamaged DNA, ζ has somewhat higher fidelity than the Y-family polymerases, but lower fidelity than the other B-family members (Fig. 2). The ability of ζ to generate both base substitutions and indels at relatively high rates is consistent with
its known participation in a large majority of spontaneous mutations, as well as in mutagenesis induced by a variety of DNA-damaging agents (Lawrence 2002). Polymerase $\zeta$’s high base substitution error rate clearly demonstrates that it has relatively low nucleotide selectivity, consistent with a possible direct role in mutagenic misinsertion of dNTPs in vivo. Polymerase $\zeta$ also efficiently extends terminal mismatches when copying undamaged DNA, as well as efficiently extending damaged termini, the latter being consistent with a role for $\zeta$ in the extension step of TLS in the two-polymerase model (Fig. 4B). A similar role has also been proposed for polymerase $\kappa$, which, like $\zeta$, is promiscuous for mismatch extension (for review, see Prakash and Prakash 2002). During DNA synthesis in vitro, $\zeta$ also generates “complex” mutations that contain multiple substitutions and indels within a short tract of DNA (Sakamoto et al. 2007; Stone et al. 2009). Consistent with this property, $\zeta$ also generates complex errors in vivo, which could be significant from an evolutionary perspective. The biological relevance of TLS is perhaps best illustrated by the role of polymerase $\nu$ in the mutagenic SOS response in E. coli, and the fact that loss of polymerase $\eta$ function in humans and in mice results in sensitivity to sunlight, predisposition to skin cancer, and altered specificity of somatic hypermutation of immunoglobulin genes. The topics and the TLS ability and fidelity of various polymerases when encountering a wide range of structurally diverse lesions have been described in great detail elsewhere (see, e.g., Prakash and Prakash 2002; Fuchs et al. 2004; Diaz and Lawrence 2005; Friedberg et al. 2005; Yang and Woodgate 2007).

**DNA Mismatch Repair**

Replication errors are corrected by DNA mismatch repair (MMR) (for review, see Kunkel and Erie 2005; Iyer et al. 2006; Hsieh and Yamane 2008). The reactions and proteins catalyzing MMR are evolutionarily conserved from *E. coli* (Fig. 5A) through humans (Fig. 5B). MMR requires initial recognition of mismatches by bacterial MutS protein or its eukaryotic homologs (Msh2-Msh6 or Msh2-Msh3). This is followed by binding of a second protein, MutL, or its eukaryotic homologs Mlh1-Pms1 (Pms2 in humans), Mlh1-Mlh2, or Mlh1-Mlh3. These MutS and MutL proteins bind and hydrolyze ATP, and in so doing, these complexes undergo conformational changes that help to coordinate the multiple protein partnerships and reactions needed to find the strand-discrimination signal, incise the nascent strand, excise the replication error, correctly synthesize new DNA, and then ligate the nascent strand. In addition to their functions in repairing replication errors, some MMR proteins also participate in other DNA transactions, including critical environmental stress-response pathways such as repair of double-stranded DNA breaks and DNA-damage surveillance to signal apoptosis. As a consequence, loss of MMR is associated with elevated mutation rates and altered survival in response to DNA damage. These in turn can give rise to microbial populations with increased fitness to survive adverse environmental conditions and to somatic cells with increased resistance to chemotherapeutic agents and increased probability of tumorigenesis. MMR proteins also prevent recombination between DNA sequences with imperfect homology, thereby influencing speciation. Some MMR proteins participate in meiotic recombination, such that loss of MMR results in infertility. MMR also modulates somatic hypermutation of immunoglobulin genes, such that loss of MMR protein functions alters the specificity of somatic hypermutation (SHM).

**Modulating MMR Efficiency**

Given these many functions and biological effects, intensive studies of MMR have revealed several ways to modulate MMR activity (Table 2). These include partial or complete inactivation by mutations in various MMR genes (for review, see Kunkel and Erie 2005; Iyer et al. 2006; Hsieh and Yamane 2008), cadmium inhibition of MMR in budding yeast (Jin et al. 2003), silencing Mlh1 expression by promoter hypermethylation (see, e.g., Herman et al. 1998), saturating MMR repair capacity under conditions of stress (Schaaper and Radman 1989), and imbalanced expression of certain MMR proteins that can reduce MMR efficiency (see Drummond et al. 1997; Shcherbakova et al. 2001 and references therein). Indeed, just as for nucleotide selectivity and proofreading, the contribution of MMR to replication fidelity can vary over a wide range (Fig. 2). On average, complete loss of bacterial MutS-dependent MMR or eukaryotic Msh2-dependent MMR elevates point mutation rates ~100-fold to 1000-fold. On the edges of the MMR efficiency continuum are some mismatches that are poorly corrected, e.g., about fivefold for 8-oxo-G•A mismatches (Pavlov et al. 2003), and others that are repaired incredibly efficiently, e.g., exceeding 10,000-fold for single-base deletion mismatches in long homo-nucleotide runs. The latter illustrate that MMR is the major guardian against the instability of repetitive sequence elements, which, as explained above, are prone to slippage and poorly proofread.

**CONCLUSIONS**

In the half century since the DNA double helix was described, we have come to more fully, but still incompletely, appreciate the elegant complexity of the DNA transactions required to replicate genomes with the fidelity needed to maintain genetic identity in the face of environmental insults, coupled with the advantage of some promiscuity for survival and evolution. This flexibility stems from the wide variability in the contributions of the three processes that determine DNA synthesis fidelity (Fig. 2) and the fact that proofreading and MMR, the two main replication-error correction mechanisms that ensure genome stability, are not essential for cellular survival. Thus, neither proofreading nor MMR contributes to replication of RNA viruses, and these viruses have very high mutation rates that fit their lifestyle (Drake and Holland 1999). Some DNA viruses use proofreading to achieve lower mutation rates but do not take advantage of their host’s MMR machinery (e.g., bacte-
riophage T4) (Santos and Drake 1994). As a consequence, they have mutation rates per base pair that are higher than organisms that do use MMR. Interestingly, given the differences in genome size, the mutation rate per genome is relatively constant among DNA-based organisms, at 0.003 (Drakes rule; Drake 1991, 1999). Also of interest is the fact that the genomes of certain bacteria do not encode obvious homologs of the major MMR genes. This leads one to wonder whether they forego MMR altogether or correct replication errors in a manner yet undiscovered.

**ACKNOWLEDGMENTS**

I thank Alan Clark and Mercedes Arana for thoughtful comments on the manuscript. The research conducted in the author’s laboratory is supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences (Projects Z01 ES065070 and Z01 ES065089).

**REFERENCES**


Evolving Views of DNA Replication (In)Fidelity

T.A. Kunkel

Access the most recent version at doi:10.1101/sqb.2009.74.027

References

This article cites 63 articles, 20 of which can be accessed free at:
http://symposium.cshlp.org/content/74/91.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Cold Spring Harbor Symposia on Quantitative Biology go to:
http://symposium.cshlp.org/subscriptions

Copyright © 2009, Cold Spring Harbor Laboratory Press