

## *Saccharomyces cerevisiae* DNA Polymerase $\delta$

HIGH FIDELITY FOR BASE SUBSTITUTIONS BUT LOWER FIDELITY FOR SINGLE- AND MULTI-BASE DELETIONS\*

Received for publication, May 12, 2005, and in revised form, June 13, 2005  
Published, JBC Papers in Press, June 17, 2005, DOI 10.1074/jbc.M505236200

John M. Fortune<sup>‡</sup>, Yuri I. Pavlov<sup>‡,§</sup>, Carrie M. Welch<sup>¶</sup>, Erik Johansson<sup>||</sup>, Peter M. J. Burgers<sup>¶</sup>,  
and Thomas A. Kunkel<sup>‡,\*\*</sup>

From the <sup>‡</sup>Laboratory of Molecular Genetics and Laboratory of Structural Biology, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina 27709, the <sup>¶</sup>Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri 63110, and the <sup>||</sup>Department of Medical Biochemistry and Biophysics, Umeå University, Umeå SE-901 87, Sweden

Eukaryotic DNA polymerase  $\delta$  (Pol  $\delta$ ) plays an essential role in replicating large nuclear genomes, a process that must be accurate to maintain stability over many generations. Based on kinetic studies of insertion of individual dNTPs opposite a template guanine, Pol  $\delta$  is believed to have high selectivity for inserting correct nucleotides. This high selectivity, in conjunction with an intrinsic 3'-exonuclease activity, implies that Pol  $\delta$  should have high base substitution fidelity. Here we demonstrate that the wild type *Saccharomyces cerevisiae* three-subunit Pol  $\delta$  does indeed have high base substitution fidelity for the 12 possible base-base mismatches, producing on average less than 1.3 stable misincorporations/100,000 nucleotides polymerized. Measurements with exonuclease-deficient Pol  $\delta$  confirm the high nucleotide selectivity of the polymerase and further indicate that proofreading enhances the base substitution fidelity of the wild type enzyme by at least 60-fold. However, Pol  $\delta$  inefficiently proofreads single nucleotide deletion mismatches in homopolymeric runs, such that the error rate is 30 single nucleotide deletions/100,000 nucleotides polymerized. Moreover, wild type Pol  $\delta$  frequently deletes larger numbers of nucleotides between distantly spaced direct repeats of three or more base pairs. Although wild type Pol  $\delta$  and Pol  $\epsilon$  both have high base substitution fidelity, Pol  $\delta$  is much less accurate than Pol  $\epsilon$  for deletions involving repetitive sequences. Thus, strand slippage during replication by wild type Pol  $\delta$  may be a primary source of insertion and deletion mutagenesis in eukaryotic genomes.

DNA polymerases are involved in replication of nuclear and mitochondrial chromosomes, DNA repair, recombination, somatic hypermutation of immunoglobulin genes, and cell cycle checkpoint control (1–3). Among these many tasks, by far the most DNA synthesis occurs during replication of large eukaryotic nuclear genomes. Three eukaryotic DNA polymerases are required for normal chromosomal DNA replication: DNA poly-

merases  $\alpha$  (Pol  $\alpha$ ),<sup>1</sup>  $\delta$  (Pol  $\delta$ ), and  $\epsilon$  (Pol  $\epsilon$ ). Pol  $\alpha$  is responsible for limited synthesis to initiate replication at origins and to prime Okazaki fragments, whereas Pol  $\delta$  is responsible for their elongation and maturation (for review, see Ref. 4). Several studies assign Pol  $\epsilon$  as the leading strand DNA polymerase, although exceptions to this rule have been documented (4).

Accurate DNA synthesis during replication is essential for maintaining genomic stability over many generations (5). The ability of Pol  $\delta$  and Pol  $\epsilon$ , which are responsible for most replication, to copy DNA accurately is enhanced by an intrinsic 3'-exonuclease activity that both enzymes possess. The importance of accurate DNA replication has prompted several studies of DNA synthesis fidelity by Pol  $\alpha$ , Pol  $\delta$ , and Pol  $\epsilon$ . Studies conducted before recombinant enzymes were available, using polymerases purified from mammalian tissues, demonstrated that the catalytic subunits of Pol  $\alpha$  (6, 7), Pol  $\epsilon$  (at that time referred to as Pol  $\delta$ -II) (8), and Pol  $\delta$  (9) can all synthesize DNA accurately. However, Pol  $\delta$  and Pol  $\epsilon$  were more accurate than Pol  $\alpha$ , and their fidelity was reduced in reactions containing high dNTP or dNMP concentrations (8, 9), conditions that partially diminish proofreading by their intrinsic 3'-exonuclease activities.

The importance of proofreading to eukaryotic replication fidelity was demonstrated further by genetic studies showing elevated mutation rates in yeast and mouse cells containing mutations that inactivate the 3'-exonuclease activity of Pol  $\delta$  (10–13) or Pol  $\epsilon$  (11). In *Saccharomyces cerevisiae*, the availability of exonuclease-deficient polymerase mutants allows for purification of both wild type and exonuclease-deficient forms of Pol  $\delta$  (14) and Pol  $\epsilon$  (15). These polymerases are multisubunit complexes that contain the polymerase/exonuclease catalytic subunits (125 kDa for Pol  $\delta$ , 256 kDa for Pol  $\epsilon$ ) physically associated with accessory subunits (55- and 40-kDa subunits for Pol  $\delta$  or 80-, 34-, and 29-kDa subunits for Pol  $\epsilon$ ). Determination of the fidelity of DNA synthesis by these exonuclease-deficient enzyme complexes defines the inherent nucleotide selectivity of each polymerase active site. These fidelity measurements can then be compared with those of wild type, exonuclease-proficient complexes to describe the apparent contribution of proofreading to replication fidelity. The contribution of proofreading deduced in this manner is more definitive than early reports for mammalian enzymes mentioned above, which relied on altered reaction conditions that only partially diminish proofreading.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>§</sup> Present address: Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198.

\*\* To whom correspondence should be addressed: NIEHS, National Institutes of Health, P. O. Box 12233, MD E3-01, Research Triangle Park, NC 27709. Tel.: 919-541-2644; Fax: 919-541-7593; E-mail: kunkel@niehs.nih.gov.

<sup>1</sup> The abbreviations used are: Pol  $\alpha$ ,  $\delta$ , and  $\epsilon$ , DNA polymerase  $\alpha$ ,  $\delta$ , and  $\epsilon$ , respectively; Exo<sup>-</sup>, exonuclease-deficient; Exo<sup>+</sup>, exonuclease-proficient; PCNA, proliferating cell nuclear antigen; wt, wild type.

Replication fidelity is usually quantified in one of two ways. The first is illustrated by several recent studies that reported steady-state kinetic parameters for Pol  $\delta$  insertion of individual correct (16–18) or incorrect (17, 18) nucleotides using an oligonucleotide primer-template. These studies indicate that *Schizosaccharomyces pombe* Pol  $\delta$  (in the presence of *Escherichia coli* single-stranded DNA-binding protein, PCNA, and replication factor C) (17) and *S. cerevisiae* Pol  $\delta$  (either in the absence or presence of PCNA) (18) both efficiently discriminate against misinsertion of dAMP, dGMP, and dTMP opposite a specific template guanine. Somewhat surprisingly, however, *S. pombe* Pol  $\delta$  did not efficiently proofread any of the three mismatches, and *S. cerevisiae* Pol  $\delta$  failed to proofread two of the three mismatches efficiently, leading to the conclusion that wild type *S. cerevisiae* Pol  $\delta$  has lower fidelity than wild type Pol  $\epsilon$  for single base misincorporations.

A second approach gives a more comprehensive view of replication fidelity. A single-stranded gapped region of M13mp2 is filled by a polymerase in the presence of all four dNTPs, and polymerase errors are scored by M13 plaque colors. This forward mutation assay permits a quantitative description of base substitution and insertion/deletion error rates at numerous template positions, representing all 12 possible base-base mismatches and a variety of different insertions and deletions in numerous sequence contexts. Using this approach, we have previously described the fidelity of the four-subunit *S. cerevisiae* Pol  $\alpha$ -primase complex (19) and the fidelity of the four-subunit, wild type and exonuclease-deficient forms of *S. cerevisiae* Pol  $\epsilon$  (20). The present study reports the fidelity of the third major replicative polymerase from *S. cerevisiae*, the three-subunit, wild type and exonuclease-deficient forms of Pol  $\delta$ . Given the inefficient proofreading of mismatches by yeast Pol  $\delta$  observed in the kinetic studies mentioned above, we investigated whether wild type yeast Pol  $\delta$  proofreads mismatched intermediates in a complete synthesis reaction that contains all four deoxynucleoside triphosphates and requires both nucleotide misinsertion and primer extension to score errors. In addition, we determined the extent to which proofreading corrects specific base-base mismatches and insertion/deletion errors in different sequence contexts. Understanding Pol  $\delta$  fidelity and proofreading ability is important because loss of Pol  $\delta$  proofreading can have serious consequences. For example, mice homozygous for exonuclease-deficient Pol  $\delta$  develop tumors at a high rate (13, 21).

#### EXPERIMENTAL PROCEDURES

**Materials**—The *S. cerevisiae* three-subunit Pol  $\delta$  was purified as described previously (14) in two forms, the wild type exonuclease-proficient enzyme complex and a complex deficient in 3'-exonuclease activity because of the D321A/E323A double amino acid replacement, designated *pol3-01* (10, 22). All materials for fidelity assays were from sources described previously (20, 23).

**Exonuclease Assays**—The 3'-exonuclease activity of the enzyme preparations was measured using a DNA template-primer formed by annealing the  $^{32}\text{P}$ -5'-end-labeled 18-mer 5'-TGA CCA TGT ACA TCA GAA-3' and the 36-mer 3'-ACT GGT ACA TGT AGT CTT AGC CTG CAT ATA GTC ACT-5'. Reaction mixtures (10  $\mu\text{l}$ ) contained 0, 1, 2, 5, or 20 nM Pol  $\delta$ , 10 nM DNA, 40 mM Tris-HCl, pH 7.8, 8 mM magnesium acetate, 75 mM NaCl, 50  $\mu\text{g/ml}$  bovine serum albumin, and either 0 or 50  $\mu\text{M}$  each dATP, dCTP, dGTP, and dTTP. Samples were incubated at 37 °C for 5 min, mixed with 10  $\mu\text{l}$  of formamide dye solution, and subjected to electrophoresis on a denaturing 7 M urea and 12% acrylamide gel in 89 mM Tris borate, pH 8.3, and 2.5 mM EDTA.

**Polymerase Fidelity Assays**—DNA synthesis fidelity was measured using several bacteriophage M13mp2 gap filling assays as described previously, but using gel-purified gapped substrates (23). Double-stranded M13mp2 DNA with a single-stranded region containing a portion of the *lacZ* gene was used as a substrate for *in vitro* DNA synthesis. For Pol  $\delta$ , reaction mixtures contained 0.15 nM DNA template, 40 mM Tris-HCl, pH 7.8, 8 mM magnesium acetate, 250  $\mu\text{M}$  dNTPs, and 9

nM exonuclease-deficient or 64 nM wild type Pol  $\delta$ . For Pol  $\epsilon$ , reaction mixtures contained 0.15 nM DNA template, 50 mM Tris-HCl, pH 7.8, 8 mM magnesium chloride, 2 mM dithiothreitol, 100  $\mu\text{g/ml}$  bovine serum albumin, 10% glycerol, 250  $\mu\text{M}$  dNTPs, and 0.6 nM exonuclease-deficient or 2.2 nM wild type Pol  $\epsilon$ . Reactions were incubated at 30 °C for up to 60 min. Polymerase concentrations and reaction times were chosen such that the entire single-stranded region of each substrate was filled by the polymerase. Reaction mixtures were analyzed by agarose gel electrophoresis to confirm that gap filling was complete. Aliquots of DNA from each reaction were introduced into *E. coli*, which were plated as described to score M13 plaque colors that reflect either correct synthesis or errors made during *in vitro* DNA synthesis (23).

The forward mutation assay utilizes a DNA substrate containing a 407-nucleotide gap with the wild type *lacZ* sequence that encodes a blue plaque phenotype. This substrate allows for detection of many different base substitution and frameshift errors that yield light blue and colorless plaques. To define specific errors, single-stranded DNA was isolated from independent mutant plaques, and the *lacZ* gene was sequenced using either an ABI Prism 377 or ABI Prism 3100 sequencer. Error rates for individual types of mutation were calculated according to Equation 1

$$\text{ER} = ((N_i/N) \times \text{MF}) / (D \times 0.6) \quad (\text{Eq. 1})$$

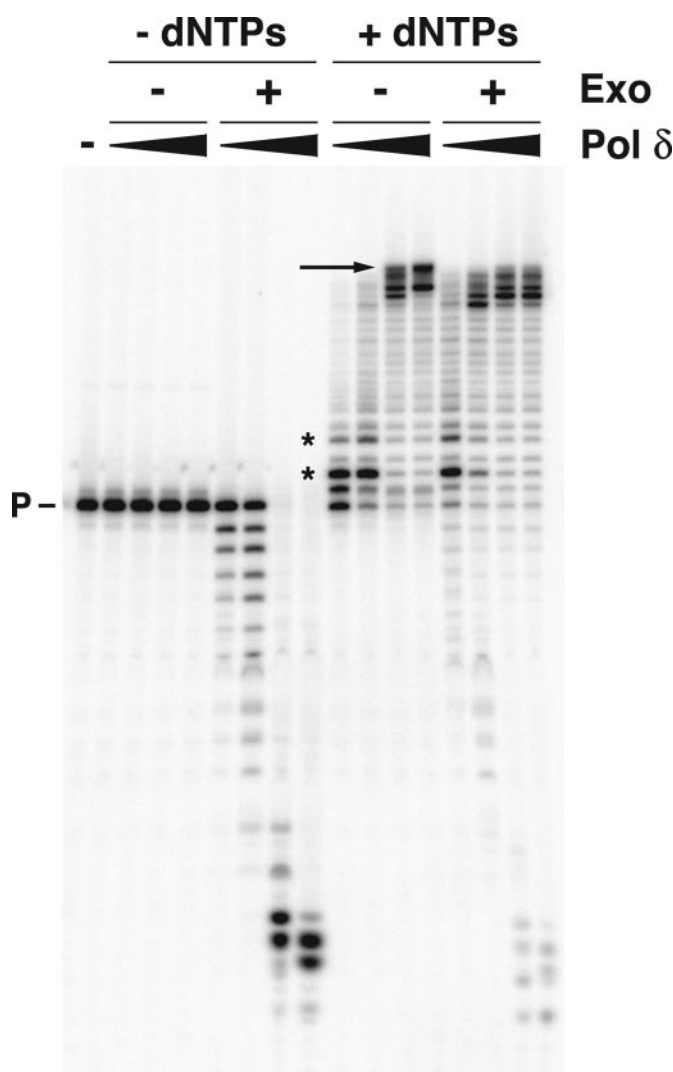
where  $N_i$  is the number of mutations of a particular type,  $N$  is the total number of mutants analyzed, MF is frequency of *lacZ* mutants,  $D$  is the number of detectable sites for the particular type of mutation, and 0.6 is the probability of expressing a mutant *lacZ* allele in *E. coli* (23).

Reversion assays were employed using substrates with similar size gaps prepared from M13mp2 phage with preexisting mutations that result in colorless plaques. With these assays, polymerase errors are scored as blue plaque revertants. Reversion substrates contained either a TGA stop codon that reverts via a polymerase-generated base substitution error, or a run of three, five, or seven T nucleotides that reverts via a single nucleotide deletion. Error rate calculations for reversion substrates were performed as above, with  $N_i/N$  representing the fraction of mutations observed within the target sequence (*i.e.* TGA codon or T nucleotide run).

#### RESULTS

**Exonuclease Activity of Wild Type and Mutant Pol  $\delta$** —The three-subunit Pol  $\delta$  was purified from *S. cerevisiae* in two forms, the wild type exonuclease-proficient enzyme complex and a complex containing a D321A/E323A double amino acid replacement in the catalytic subunit (10, 14, 22). This double mutation, corresponding to the *pol3-01* allele, previously has been shown to inactivate 3'-exonuclease function (12, 18, 22). To assess the exonuclease activity of the current wild type and *pol3-01* enzyme preparations, their ability to degrade the primer strand of an oligonucleotide template-primer was examined. As expected, the 3'-exonuclease of wild type yeast Pol  $\delta$  degraded the primer in the absence of dNTPs (Fig. 1). In the presence of dNTPs, a small amount of exonucleolytic degradation by the wild type enzyme was observed, but the major products, including a full-length 36-mer, resulted from polymerization (Fig. 1). In contrast, no exonuclease activity was observed for the *pol3-01* enzyme preparation (Fig. 1). This lack of exonuclease activity confirms that the *pol3-01* mutant enzyme preparation used here is exonuclease-deficient. The *pol3-01* enzyme was active as a polymerase in the presence of dNTPs and even generated a product that was one nucleotide longer than the template (*arrow* in Fig. 1), an end-addition product characteristic of exonuclease-deficient DNA polymerases (24, 25). Also notable are the variable band intensities of different chain lengths (*e.g.* compare the bands with *asterisks* with their neighbors in Fig. 1), indicating that the wild type and exonuclease-deficient Pol  $\delta$  catalytic subunits both have difficulty replicating certain template sequences despite the presence of the p55 and p40 accessory subunits.

**Fidelity of Wild Type and Exonuclease-deficient Pol  $\delta$** —Previous studies reported steady-state kinetic parameters only for insertion of a single correct or incorrect nucleotide opposite a single template guanine by Pol  $\delta$  (17, 18). Therefore, to obtain



**FIG. 1. Exonuclease activity of wild type and mutant Pol  $\delta$ .** Exonuclease activity of the wild type *S. cerevisiae* Pol  $\delta$  (Exo+) was compared with that of the *pol3-01* mutant (Exo-). The DNA substrate was an oligonucleotide-based template-primer that contained a  $^{32}\text{P}$ -5'-end-labeled primer strand. The position of the  $^{32}\text{P}$ -labeled primer on a polyacrylamide gel is indicated (P). Reactions contained 10 nM DNA and 0, 1, 2, 5, or 20 nM Pol  $\delta$ , either in the presence or absence of 50  $\mu\text{M}$  dNTPs. DNA extension products of varying intensities are marked by asterisks, and end-addition products generated by exonuclease-deficient Pol  $\delta$  are marked with an arrow.

a comprehensive view of rates for a variety of substitution and insertion/deletion errors in numerous sequence contexts, we determined the fidelity of wild type and exonuclease-deficient Pol  $\delta$  during gap filling DNA synthesis in the M13mp2 forward mutation assay. Both forms of Pol  $\delta$  completely copied the *lacZ* template sequence within a 407-nucleotide single-stranded DNA gap (data not shown, but indistinguishable from Fig. 3 in Ref. 23). When the products of these reactions were introduced into *E. coli* cells, *lacZ* mutant frequencies among the M13mp2 plaques were 0.0047 for wild type Pol  $\delta$  and 0.023 for exonuclease-deficient Pol  $\delta$  (Table I, average of duplicate measurements). Mutant plaques were selected, and their DNA was sequenced to identify the nucleotide changes responsible for the phenotypes (Fig. 2). Using this information, the rates for specific classes of errors were calculated (see "Experimental Procedures"), and error rates for the two enzymes were compared to establish the contribution of nucleotide selectivity and proofreading to fidelity (Table I and described below).

**Base Substitutions**—The mutations identified in DNA prod-

ucts generated by wild type Pol  $\delta$  included only nine base substitutions (Table I), such that the substitution frequency was not significantly higher than the background of the assay (see Footnote *a* of Table I). The average error rate is  $\leq 1.3 \times 10^{-5}$ , i.e. wild type Pol  $\delta$  generates less than one single base substitution error for every 80,000 nucleotides polymerized. In contrast, exonuclease-deficient Pol  $\delta$  generated base substitutions at many different template positions (Fig. 2), reflecting all possible mismatches except for C-dCTP. The average base substitution error rate is  $13 \times 10^{-5}$ , at least 10-fold higher than for wild type Pol  $\delta$ . Error rates were also calculated individually for the 12 different mismatches, each of which can be scored at between 9 and 27 template nucleotides in the *lacZ* mutational target. These rates varied over more than a 12-fold range, from the highest rate of  $14 \times 10^{-5}$  for T-dGTP to the lowest rate of  $\leq 1.1 \times 10^{-5}$  for C-dCTP (Fig. 3). Comparisons with previous results for yeast Pol  $\alpha$  (19), which is naturally exonuclease-deficient, and an exonuclease-deficient derivative of yeast Pol  $\epsilon$  (20) reveal that base substitution error rates of these three replicative polymerases differ by less than 2-fold for most mismatches but that error rate differences of up to 20-fold exist (e.g. T-dTTP,  $1.2 \times 10^{-5}$  for Pol  $\delta$  versus  $24 \times 10^{-5}$  for Pol  $\epsilon$ ).

**Small Insertions and Deletions**—The mutations identified in the DNA products generated by wild type Pol  $\delta$  included 14 single nucleotide deletions (Table I), resulting in a specific mutant frequency ( $1.6 \times 10^{-3}$ ) that is 30-fold higher than the background of the assay (see Footnote *a* of Table I) and yielding a wild type Pol  $\delta$  error rate of  $1.3 \times 10^{-5}$ . Exonuclease-deficient Pol  $\delta$  generated 119 single nucleotide deletions (Table I), yielding an average error rate of  $5.7 \times 10^{-5}$ . This is only 4-fold higher than the error rate for wild type Pol  $\delta$ , suggesting that perhaps 25% of single nucleotide deletion mismatches escape proofreading. Exonuclease-deficient Pol  $\delta$  also generated single nucleotide additions and two-nucleotide deletions at rates of  $1.2 \times 10^{-5}$  and  $1.0 \times 10^{-5}$ , respectively. The wild type Pol  $\delta$  rates for these errors ( $\leq 0.09 \times 10^{-5}$ ) were at least 11-fold lower, suggesting that more than 90% of the insertion/deletion mismatches responsible for these errors are proofread.

**Larger Deletions**—Also recovered from the DNA products generated by wild type Pol  $\delta$  were 18 *lacZ* mutants missing larger numbers of nucleotides (Table II). All 18 involved deletion of one copy of a 5-base or 6-base direct repeat sequence plus the intervening bases, and no such deletions can be attributed to the background mutant frequency (26). Exonuclease-deficient Pol  $\delta$  generated 57 large deletions, 48 of which involved deletion of one copy of a direct repeat sequence plus the intervening bases. All observed deletions between direct repeats are predicted to involve a looped intermediate that is stabilized by three to eight base pairs (see Footnote *b*, Table II) (27). The direct repeat deletion frequency for exonuclease-deficient Pol  $\delta$  was only 1.4-fold higher than for wild type Pol  $\delta$  (0.0027 versus 0.0020, Table II), suggesting that proofreading does not efficiently correct misaligned intermediates that yield large deletions between direct repeats.

**Comparative Proofreading Efficiencies of Pol  $\delta$  and Pol  $\epsilon$** —A previous kinetic study of nucleotide insertions opposite a single defined template guanine concluded that the fidelity of wild type *S. cerevisiae* Pol  $\delta$  is significantly lower than that of Pol  $\epsilon$  and that this is the result of inefficient proofreading by wild type Pol  $\delta$  (18). To investigate this hypothesis further, we evaluated the base substitution and deletion fidelity of Pol  $\delta$  and Pol  $\epsilon$  in complete synthesis reactions where all four dNTPs are in competition. Both wild type and exonuclease-deficient forms of Pol  $\delta$  and Pol  $\epsilon$  were tested, allowing a direct comparison of proofreading efficiencies for the two polymerases. Fidelity determinations were made using reversion assays. These

TABLE I  
Error rates for Pol  $\delta$  base substitutions and frameshifts in a forward mutation assay

Error	Pol $\delta$ (Exo-)		Pol $\delta$ (wt)		Exo-/Exo+
	No.	Error rate <sup>a</sup> ( $\times 10^{-5}$ )	No.	Error rate <sup>a</sup> ( $\times 10^{-5}$ )	
Base substitutions	171	13	9	$\leq 1.3$	$\geq 10$
-1 deletions	119	5.7	14	1.3	4.4
+1 insertions	25	1.2	0	$< 0.094$	$\geq 13$
-2 deletions <sup>b</sup>	21	1.0	1	0.094	11
Other errors <sup>c</sup>	67	380 (MF)	18	200 (MF)	
Total	403		42		
Mutant frequency <sup>a</sup>	0.023		0.0047		

<sup>a</sup> Error rates and mutant frequencies were calculated using data from two independent experiments. Background mutant frequencies in the forward mutation assay are  $3.4 \times 10^{-4}$  for base substitutions and  $0.55 \times 10^{-4}$  for single nucleotide insertions/deletions (26). From these frequencies, background error rates are  $0.45 \times 10^{-5}$  and  $0.046 \times 10^{-5}$ , respectively.

<sup>b</sup> In error rate calculations for -2 deletions, the number of detectable sites in the *lacZ* target sequence was assumed to be the same as for -1 deletions (23).

<sup>c</sup> Other errors for Pol  $\delta$  (Exo-) included 53 large deletions, 48 of which were deletions between direct repeat sequences (see Table II), five complex events involving multiple large deletions, six tandem frameshift/base substitution mutations, and three -3 frameshift events. Other errors for Pol  $\delta$  (wt) were all large deletions between direct repeats (Table II). The likelihood of other errors being generated is expressed as a mutant frequency (MF) rather than an error rate.

assays score polymerase errors as blue plaque revertants of M13mp2 derivatives that normally have colorless plaque phenotypes caused by preexisting substitution or frameshift mutations in the *lacZ* gene. Because reversion assays score substitution or insertion/deletion errors at many fewer sites than does the forward mutation assay described above, they have lower background mutant frequencies and thus provide more sensitive determinations of error rates (28). This is especially critical when evaluating proofreading-proficient polymerases that are often highly accurate.

Three assays were used to monitor reversion frequencies for products of gap filling synthesis by the multisubunit, wild type and exonuclease-deficient forms of Pol  $\delta$  and Pol  $\epsilon$ . The first assay scores single base substitutions that revert a TGA codon at *lacZ* nucleotides 87-89. Synthesis by both wild type polymerases resulted in reversion frequencies that were similar to the background reversion frequency (Table III, first line). The average error rate for the eight single base mismatches that can be scored at the TGA stop codon is therefore  $\leq 1.3 \times 10^{-5}$  for wild type Pol  $\delta$  and  $\leq 1.8 \times 10^{-5}$  for wild type Pol  $\epsilon$  (Table IV). Average error rates for the exonuclease-deficient polymerases are much higher, indicating that proofreading contributes to the single base substitution fidelity of Pol  $\delta$  and Pol  $\epsilon$  by factors of at least 60-fold and 24-fold, respectively (Table IV). Thus, both Pol  $\delta$  and Pol  $\epsilon$  were found to be highly accurate for base substitutions, and proofreading by both enzymes contributes to this high fidelity.

Two additional reversion assays were employed to score insertion and deletion errors that revert a +1 frameshift mutation in *lacZ* within a homopolymeric run of either five or seven template thymines. With the DNA substrate containing five consecutive thymines, synthesis by wild type Pol  $\delta$  resulted in a reversion frequency of  $66 \times 10^{-5}$  (Table III, second line). This value reflects polymerization errors *in vitro* because it is well above the background reversion frequency. In fact, the reversion frequency for wild type Pol  $\delta$  is only 3.5-fold lower than that observed for exonuclease-deficient Pol  $\delta$ . With the DNA substrate containing seven consecutive thymines, synthesis by wild type Pol  $\delta$  resulted in a reversion frequency of  $170 \times 10^{-5}$  (Table III, third line), a value that is only 1.5-fold lower than that observed for exonuclease-deficient Pol  $\delta$ . The results with Pol  $\epsilon$  were different. Reversion frequencies for wild type Pol  $\epsilon$  were lower than its exonuclease-deficient derivative by factors of 5-fold and 11-fold, respectively, for the T<sub>5</sub> and T<sub>7</sub> runs (Table III).

With both substrates containing homopolymeric thymine runs, reversion could result from either single nucleotide deletions or two-nucleotide additions within the runs, or from de-

letions of single nucleotides flanking the runs. To distinguish between these mutagenic events, we sequenced independent revertants generated by wild type and exonuclease-deficient forms of Pol  $\delta$  and Pol  $\epsilon$ . The predominant error was found to be deletion of a single thymine in the run. From this sequencing information and the reversion frequencies in Table III, we calculated single nucleotide deletion error rates for all four polymerases with the seven-thymine run (Table IV). Error rates reveal that wild type Pol  $\epsilon$  is 13-fold more accurate than wild type Pol  $\delta$ . Furthermore, Pol  $\epsilon$  proofreads mismatches with a single unpaired template thymine in the run much more efficiently than does Pol  $\delta$ .

#### DISCUSSION

This study completes a comprehensive analysis of the base substitution and insertion/deletion fidelity of multisubunit forms of Pol  $\alpha$  (19), Pol  $\delta$  (this study), and Pol  $\epsilon$  (20), the three major replicative DNA polymerases in *S. cerevisiae*. The results have several implications regarding polymerase nucleotide selectivity and proofreading efficiency, not only for single base substitutions, but also for small deletions and additions and for deletions of multiple bases between direct repeats.

*Nucleotide Selectivity without Proofreading*—The nucleotide selectivity of these three polymerase complexes, *i.e.* their ability to select correct nucleotides for incorporation into properly aligned primer-templates, can be considered by comparing error rates for Pol  $\alpha$ , which is naturally exonuclease-deficient, with error rates for exonuclease-deficient mutant derivatives of Pol  $\delta$  and Pol  $\epsilon$ . Average error rates for the 241 single base substitutions that can be scored in the forward mutation assay are similar:  $11 \times 10^{-5}$  for the four-subunit Pol  $\alpha$ -primase complex (19),  $13 \times 10^{-5}$  for the three-subunit Pol  $\delta$  complex (Table I), and  $24 \times 10^{-5}$  for the four-subunit Pol  $\epsilon$  complex (20). For the majority of the 12 possible base-base mismatches, error rates of misinsertion by these three DNA polymerases were within 2-3-fold of each other (Fig. 3). Likewise, the average error rates among the three polymerases for the 199 single nucleotide deletions that can be scored in the forward mutation assay are remarkably similar:  $3.1 \times 10^{-5}$  for the four-subunit Pol  $\alpha$ -primase complex (19),  $5.7 \times 10^{-5}$  for the three-subunit Pol  $\delta$  complex (Table I), and  $5.6 \times 10^{-5}$  for the four-subunit Pol  $\epsilon$  complex (20). The data suggest that the active sites of these homologous B family enzymes have similar average nucleotide selectivity. Nonetheless, the polymerase-substrate interactions relevant to fidelity are not likely to be identical because the polymerases generate certain errors at substantially different rates. For example, Pol  $\epsilon$  is 20-fold less accurate than Pol  $\delta$  for substitutions resulting from T-dTTP mismatches (Fig. 3). Such

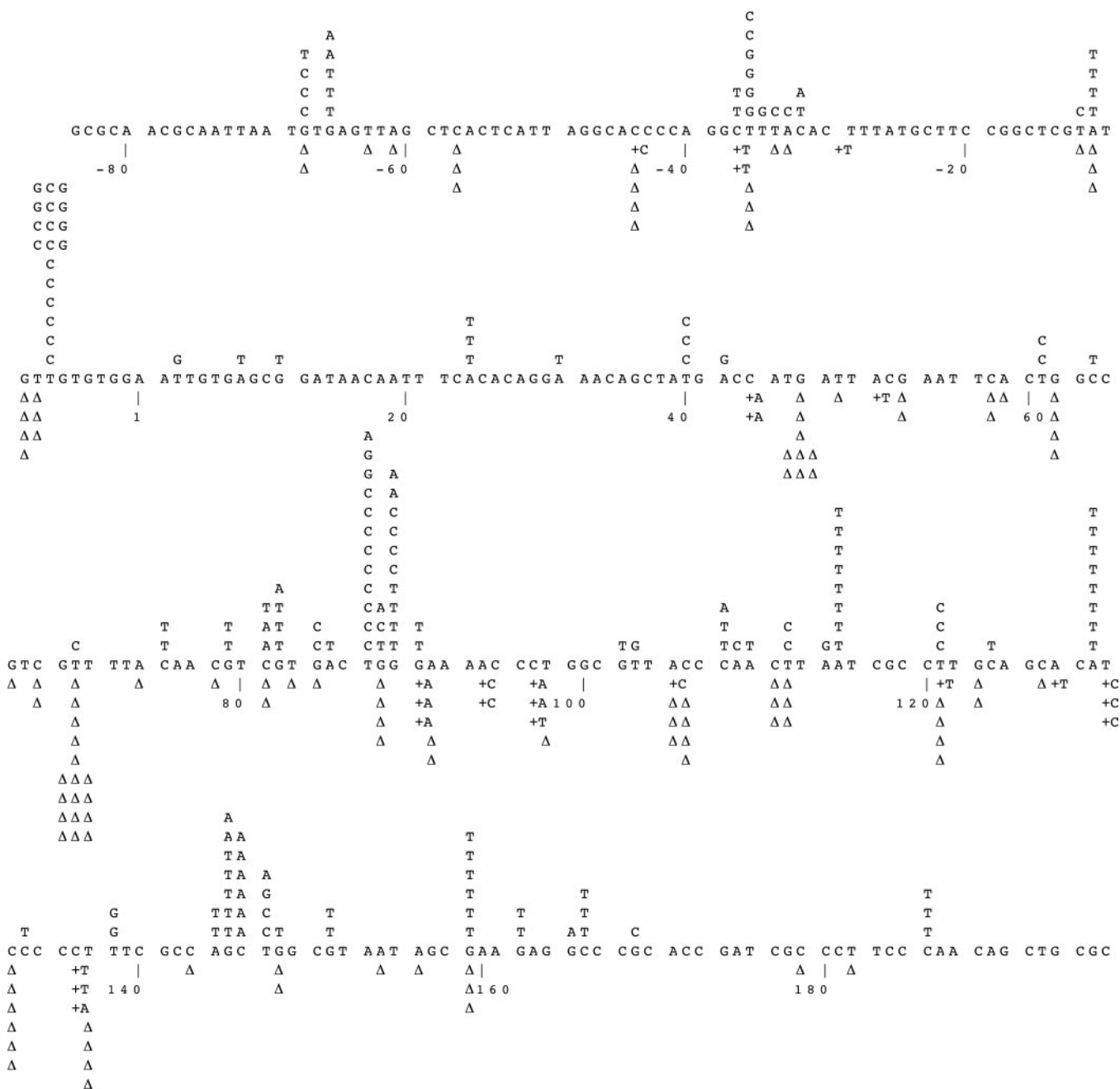


FIG. 2. **Spectrum of single base errors by exonuclease-deficient Pol  $\delta$ .** A DNA synthesis fidelity assay was performed with gapped M13mp2 substrate that contained a portion of the *lacZ* gene in the 407-nucleotide gap region. Errors generated by exonuclease-deficient Pol  $\delta$  in the M13mp2 *lacZ* target sequence are shown, with base substitution mutations above the sequence and single nucleotide deletions ( $\Delta$ ) and additions below. The *lacZ*(+) strand is shown, with the transcriptional start site designated as position 1 and the first 53 codons displayed as triplets. The gapped region ends at position 191.

polymerase-specific differences for certain errors may be useful mutational signatures in future studies that probe the roles of Pol  $\alpha$ , Pol  $\delta$ , and Pol  $\epsilon$  during leading and lagging strand chromosomal replication in yeast (29).

Although the explanation for these differences is currently unknown, we noted previously that Pol  $\epsilon$  has a 66-amino-acid insertion not found in Pol  $\alpha$  or Pol  $\delta$  which might influence base selection (see "Discussion" in Ref. 20). Moreover, the higher fidelity of Pol  $\epsilon$  for single nucleotide deletions in the  $T_7$  runs and deletions between direct repeats correlates with the higher processivity of Pol  $\epsilon$  compared with Pol  $\delta$  (14, 30). The correlation between processivity and single nucleotide deletion fidelity, extended by this study to larger deletions between direct repeats, is consistent with several earlier observations with

other polymerases (for review, see Ref. 31) and suggests that DNA strand misalignments may form during polymerase dissociation-reassociation with primer-templates (6). We have now initiated studies using M13mp2-based fidelity assays to investigate whether accessory proteins that increase the processivity of Pol  $\delta$  (e.g. PCNA) alter fidelity for either base substitutions or insertions/deletions.

**Proofreading Efficiency**—As mentioned above, prior kinetic studies presented evidence that wild type Pol  $\delta$  enzymes from budding and fission yeast do not efficiently proofread certain mismatches (17, 18). The present work extends those studies to a wide variety of base substitution and insertion/deletion errors and leads us to conclude that proofreading does indeed contribute strongly to fidelity but that the efficiency with which wild

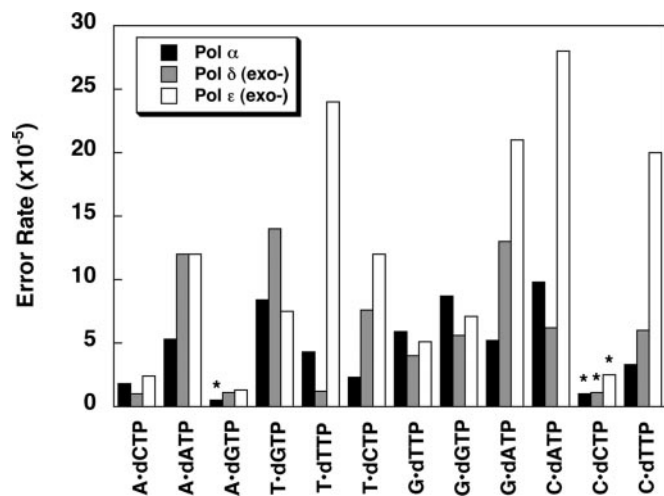


FIG. 3. **Base substitution error rates of exonuclease-deficient Pol  $\delta$  and Pol  $\epsilon$ .** Individual base-base mispair error rates are shown for Pol  $\alpha$  (black bars), exonuclease-deficient Pol  $\delta$  (gray bars), and exonuclease-deficient Pol  $\epsilon$  (open bars). Some error rates are reported as less than or equal to values (indicated by an asterisk above the bar). All data are from the forward mutation assay. Data for Pol  $\delta$  are presented in Fig. 2. Data for Pol  $\alpha$  and Pol  $\epsilon$  are from Refs. 19 and 20, respectively. Error rates for Pol  $\alpha$  have been recalculated from Ref. 19 because the number of detectable sites known to be present in the *lacZ* target sequence has increased since the time of the original work (23).

TABLE II  
Deletions between direct repeats generated by Pol  $\delta$

Mutation <sup>a</sup>	Direct repeat <sup>b</sup>	No. observed	
		Pol $\delta$ (Exo-)	Pol $\delta$ (wt)
$\Delta$ 317	CCCCG (8)	27	16
$\Delta$ 325	CTGGCG (7)	5	1
$\Delta$ 357	CCCAA (6)	1	
$\Delta$ 37	GAA (6)	1	
$\Delta$ 388	GCCC (5)	1	
$\Delta$ 138	CACCC (5)		1
$\Delta$ 315	CCCT (4)	5	
$\Delta$ 55	TTT (4)	1	
$\Delta$ 331	CCC (3)	1	
$\Delta$ 181	GGC (3)	1	
$\Delta$ 139	CCC (3)	1	
$\Delta$ 79	TF (3)	1	
$\Delta$ 59	GCT (3)	1	
$\Delta$ 46	CCC (3)	1	
$\Delta$ 24	GCC (3)	1	
Total		48	18
Mutant frequency <sup>c</sup>		0.0027	0.0020

<sup>a</sup> Mutation nomenclature is based on the number of bases deleted by Pol  $\delta$  when copying the *lacZ* target sequence.

<sup>b</sup> Deletions occur between two identical copies of the indicated sequence. Deletion events have been proposed to involve a looped intermediate, which is stabilized by both the newly synthesized strand complementary to the repeat sequence and base pairing between terminal nucleotides within the loop (27). The total number of such stabilizing base-base interactions is indicated in parentheses.

<sup>c</sup> The background mutant frequency for large deletions in the forward mutation assay is 0.00016. Almost all of these events are caused by recombination in *E. coli* rather than deletions between direct repeats (26).

type *S. cerevisiae* Pol  $\delta$  proofreads errors is highly dependent on the pre-mutational intermediate to be corrected. For base substitution errors, a 60-fold difference was observed between error rates for wild type and exonuclease-deficient Pol  $\delta$  in the TGA reversion assay (Table IV), and a greater than 10-fold difference was detected in the forward assay (Table I). Both observations demonstrate efficient proofreading of single base-base mismatches generated during a complete DNA synthesis reaction containing all four dNTPs. The biochemical data are consistent with genetic studies showing elevated mutation

TABLE III  
Fidelity of Pol  $\delta$  and Pol  $\epsilon$  in base substitution and frameshift reversion assays

DNA substrate	Reversion frequency <sup>a</sup> ( $\times 10^{-5}$ )		Reversion frequency <sup>a</sup> ( $\times 10^{-5}$ )	
	Pol $\delta$ (Exo-)	Pol $\delta$ (wt)	Pol $\epsilon$ (Exo-)	Pol $\epsilon$ (wt)
TGA <sub>89</sub>	140	$\leq 2.3$	78	$\leq 3.3$
T <sub>5</sub>	220	66	57	11
T <sub>7</sub>	250	170	200	18

<sup>a</sup> Reversion frequencies were calculated using data from two independent experiments. Background reversion frequencies in this assay are  $0.5 - 2.5 \times 10^{-5}$  (23, 28).

rates in yeast and mouse cells containing mutations that inactivate the 3'-exonuclease activity of Pol  $\delta$  (10–13). Although earlier studies led to the conclusion that wild type *S. cerevisiae* Pol  $\epsilon$  (32) has higher single base misincorporation fidelity than does wild type Pol  $\delta$  (18), our results neither support nor contradict this conclusion. Both wild type enzyme complexes are so accurate that we do not detect substitution errors above the background noise of fidelity assays. Thus, Pol  $\delta$  and Pol  $\epsilon$  have high base substitution fidelity, which reflects both high nucleotide selectivity and efficient proofreading.

In contrast, Pol  $\delta$  is less accurate for single nucleotide deletions. Error rates for exonuclease-deficient Pol  $\delta$  in the forward assay are only 4-fold higher than those for the wild type enzyme (Table I), demonstrating a deficiency in proofreading of single nucleotide deletions by Pol  $\delta$ . Interestingly, the error rate for single nucleotide additions is at least 13-fold higher for exonuclease-deficient Pol  $\delta$  compared with that for wild type polymerase, and addition errors by wild type Pol  $\delta$  are not detected above background. Thus, Pol  $\delta$  is able to distinguish between the presence of an unpaired base in the primer strand *versus* the template strand of the duplex upstream of the active site, being able to proofread the former more efficiently than the latter.

Wild type Pol  $\delta$  has high fidelity for two-nucleotide ( $-2$ ) deletions. Exonuclease-deficient Pol  $\delta$  generates  $-2$  deletions at a readily detectable rate, at least 11-fold higher than that observed for wild type  $\delta$  (Table I), which does not generate these errors at levels above background. Formation of  $-2$  deletions requires polymerization from mismatched primer-templates containing two unpaired bases in the template strand. Because the  $-2$  errors observed here do not occur in repetitive sequences, the two unpaired bases may not be stabilized by extensive correct base pairing and may therefore be near the primer terminus. The greater geometric distortion caused by two unpaired bases likely explains the efficient proofreading of  $-2$  deletion intermediates by Pol  $\delta$ .

The contribution of proofreading to the fidelity of wild type *S. cerevisiae* Pol  $\delta$  is low for single nucleotide deletions in homopolymeric runs (Table III). The misaligned primer-templates inferred to result in single nucleotide deletions contain unpaired bases that could potentially slow polymerization and trigger excision of the nascent strand by the exonuclease. However, in repetitive sequence contexts, the unpaired bases would not necessarily be at the primer terminus but could instead be present upstream of the polymerase active site and be stabilized by correct base pairs whose number would depend on the length of the homopolymeric run (33). The potential for the mismatch to escape the polymerase active site can explain why Pol  $\delta$  proofreading of single nucleotide deletions in T<sub>5</sub> and T<sub>7</sub> runs (Tables III and IV) is inefficient. In both cases, the primer-template could theoretically contain four or more correct T-A base pairs between the active site and the unpaired base. Our data suggest that this degree of separation allows Pol  $\delta$  to continue extension with only limited proofreading of the dele-

TABLE IV  
Fidelity for base substitutions, frameshifts, and deletions between direct repeats

Error type	Error rate <sup>a</sup> ( $\times 10^{-5}$ )		Pol $\delta$ Exo-/Exo+	Error rate <sup>a</sup> ( $\times 10^{-5}$ )		Pol $\epsilon$ Exo-/Exo+
	Pol $\delta$ (Exo-)	Pol $\delta$ (wt)		Pol $\epsilon$ (Exo-)	Pol $\epsilon$ (wt)	
TGA <sub>89</sub> reversion	78	$\leq 1.3$	$\geq 60$	43	$\leq 1.8$	$\geq 24$
T <sub>7</sub> reversion	40	31	1.3	33	2.3	14
$\Delta_{DR}$ (MF) <sup>b</sup>	270	200	1.4	41	$\leq 6.7$	$\geq 6.1$

<sup>a</sup> Error rates for Pol  $\delta$  and Pol  $\epsilon$  were calculated from the TGA<sub>89</sub> and T<sub>7</sub> reversion frequencies reported in Table III, as described under "Experimental Procedures."

<sup>b</sup> Because there is no defined number of detectable sites in the *lacZ* target sequence for deletions between direct repeats, the probability of this error type is expressed as a mutant frequency rather than an error rate. Mutant frequencies for Pol  $\delta$  are from Table II. Mutant frequencies for Pol  $\epsilon$  are from data in Ref. 20.

tion mismatch. These findings are consistent with *S. cerevisiae* studies *in vivo*, which show that the contribution of proofreading to replication fidelity in homopolymeric adenine runs decreases as the run length increases (34).

This logic can be extended to the inefficient proofreading of intermediates required for larger deletions between direct repeats. These deletions can readily be explained by a model in which the first repeat sequence encountered is copied, followed by primer relocation to the second repeat sequence and continued synthesis from the misaligned intermediate that contains a loop of unpaired, template strand bases (27). The preponderance of Pol  $\delta$  deletions involving loss of 317 bases including one copy of the CCCGC repeat (Table II, top line) may reflect the fact that the repeat is G-C-rich and that three additional G-C base pairs can form in the looped template strand, which would stabilize the misaligned intermediate (see Fig. 3 in Ref. 27). The stability of the misaligned intermediate imparted by as many as eight correct G-C base pairs could explain why wild type and exonuclease-deficient Pol  $\delta$  generate this particular large deletion with very high and similar frequencies ( $1.8 \times 10^{-3}$  and  $1.5 \times 10^{-3}$ , respectively). In fact, all 18 large deletions made by wild type Pol  $\delta$  involved direct repeats of at least five base pairs, indicating that its intrinsic exonuclease is largely unable to proofread deletion intermediates stabilized by five or more correct base pairs. However, even when stabilizing base pairs in both the direct repeat and looped template strand are considered (see Footnote b, Table II), 13 deletions generated by exonuclease-deficient Pol  $\delta$  are inferred to involve fewer than five correct base pairs (frequency  $0.7 \times 10^{-3}$ ). No such deletions were observed for wild type Pol  $\delta$  (frequency  $\leq 0.11 \times 10^{-3}$ ). This difference is greater than 6-fold and suggests that Pol  $\delta$  can proofread deletion intermediates stabilized by fewer than five base pairs.

The protection against proofreading afforded by correct base pairing in misaligned intermediates is consistent with structural information on the Pol  $\delta$  homolog RB69 Pol (35), which reveals protein interactions with four to five base pairs of duplex DNA upstream of the active site. However, polymerase-substrate interactions that are relevant to insertion/deletion fidelity are likely not identical among eukaryotic family B members. Exonuclease-deficient Pol  $\epsilon$  clearly generates certain deletions at lower rates than does exonuclease-deficient Pol  $\delta$ , and wild type Pol  $\epsilon$  proofreads deletion intermediates more efficiently than does wild type Pol  $\delta$  (Table IV). The combination of these differences in nucleotide selectivity and proofreading efficiency results in large differences in the insertion/deletion fidelity of these two major eukaryotic replicative polymerases. Thus, wild type Pol  $\epsilon$  is at least 30-fold more accurate than wild type Pol  $\delta$  for large deletions and 13-fold more accurate for single nucleotide deletions in the T<sub>7</sub> run (Table IV). To the extent that these differences in fidelity are maintained when all required accessory proteins are present, the differences have important cellular implications. Because these two polymerases perform the bulk of chain elongation

during DNA replication, the lower insertion/deletion fidelity of Pol  $\delta$  compared with Pol  $\epsilon$  implies that insertion and deletion errors generated by Pol  $\delta$  may be a primary source of eukaryotic genome instability, as exemplified by the renowned microsatellite instability in cells deficient in DNA mismatch repair (36).

Given the selective differences in Pol  $\delta$  and Pol  $\epsilon$  proofreading efficiency reported here, it will be interesting to explore the biological consequences of loss of proofreading activity. Studies with mice that are homozygous for an exonuclease-deficient form of Pol  $\delta$  indicate that the animals prematurely develop certain types of cancer (13, 21). It is currently not known whether loss of Pol  $\epsilon$  proofreading might have a similar effect. Future studies on Pol  $\delta$  and Pol  $\epsilon$  fidelity should provide additional insight into the roles of these essential replication proteins.

*Acknowledgments*—We thank Zachary Pursell and Stephanie Nick McElhinny for insightful comments on the manuscript.

#### REFERENCES

- Hubscher, U., Maga, G., and Spadari, S. (2002) *Annu. Rev. Biochem.* **71**, 133–163
- Shcherbakova, P. V., Bebenek, K., and Kunkel, T. A. (2003) *Sci. Aging Knowl. Environ.* RE3
- Bebenek, K., and Kunkel, T. A. (2004) *Adv. Protein Chem.* **69**, 137–165
- Garg, P., and Burgers, P. M. (2005) *Crit. Rev. Biochem. Mol. Biol.* **40**, 115–128
- Kunkel, T. A. (2004) *J. Biol. Chem.* **279**, 16895–16898
- Kunkel, T. A. (1986) *J. Biol. Chem.* **261**, 13581–13587
- Kunkel, T. A., and Alexander, P. S. (1986) *J. Biol. Chem.* **261**, 160–166
- Kunkel, T. A., Sabatino, R. D., and Bambara, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4865–4869
- Thomas, D. C., Roberts, J. D., Sabatino, R. D., Myers, T. W., Tan, C. K., Downey, K. M., So, A. G., Bambara, R. A., and Kunkel, T. A. (1991) *Biochemistry* **30**, 11751–11759
- Morrison, A., Johnson, A. L., Johnston, L. H., and Sugino, A. (1993) *EMBO J.* **12**, 1467–1473
- Morrison, A., and Sugino, A. (1994) *Mol. Gen. Genet.* **242**, 289–296
- Jin, Y. H., Obert, R., Burgers, P. M., Kunkel, T. A., Resnick, M. A., and Gordenin, D. A. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5122–5127
- Goldsby, R. E., Hays, L. E., Chen, X., Olmsted, E. A., Slayton, W. B., Spangrude, G. J., and Preston, B. D. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15560–15565
- Burgers, P. M., and Gerik, K. J. (1998) *J. Biol. Chem.* **273**, 19756–19762
- Chilkova, O., Jonsson, B. H., and Johansson, E. (2003) *J. Biol. Chem.* **278**, 14082–14086
- Einolf, H. J., and Guengerich, F. P. (2000) *J. Biol. Chem.* **275**, 16316–16322
- Chen, X., Zuo, S., Kelman, Z., O'Donnell, M., Hurwitz, J., and Goodman, M. F. (2000) *J. Biol. Chem.* **275**, 17677–17682
- Hashimoto, K., Shimizu, K., Nakashima, N., and Sugino, A. (2003) *Biochemistry* **42**, 14207–14213
- Kunkel, T. A., Hamatake, R. K., Motto-Fox, J., Fitzgerald, M. P., and Sugino, A. (1989) *Mol. Cell. Biol.* **9**, 4447–4458
- Shcherbakova, P. V., Pavlov, Y. I., Chilkova, O., Rogozin, I. B., Johansson, E., and Kunkel, T. A. (2003) *J. Biol. Chem.* **278**, 43770–43780
- Goldsby, R. E., Lawrence, N. A., Hays, L. E., Olmsted, E. A., Chen, X., Singh, M., and Preston, B. D. (2001) *Nat. Med.* **7**, 638–639
- Simon, M., Giot, L., and Faye, G. (1991) *EMBO J.* **10**, 2165–2170
- Bebenek, K., and Kunkel, T. A. (1995) *Methods Enzymol.* **262**, 217–232
- Clark, J. M., Joyce, C. M., and Beardsley, G. P. (1987) *J. Mol. Biol.* **198**, 123–127
- Clark, J. M. (1988) *Nucleic Acids Res.* **16**, 9677–9686
- Kunkel, T. A. (1985) *J. Biol. Chem.* **260**, 5787–5796
- Kunkel, T. A., and Soni, A. (1988) *J. Biol. Chem.* **263**, 14784–14789
- Kunkel, T. A., Patel, S. S., and Johnson, K. A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 6830–6834
- Pavlov, Y. I., Newlon, C. S., and Kunkel, T. A. (2002) *Mol. Cell* **10**, 207–213
- Weiser, T., Gassmann, M., Thommes, P., Ferrari, E., Hafkemeyer, P., and

- Hubscher, U. (1991) *J. Biol. Chem.* **266**, 10420–10428
31. Bebenek, K., and Kunkel, T. A. (2000) *Cold Spring Harbor Symp. Quant. Biol.* **65**, 81–91
32. Shimizu, K., Hashimoto, K., Kirchner, J. M., Nakai, W., Nishikawa, H., Resnick, M. A., and Sugino, A. (2002) *J. Biol. Chem.* **277**, 37422–37429
33. Kroutil, L. C., Register, K., Bebenek, K., and Kunkel, T. A. (1996) *Biochemistry* **35**, 1046–1053
34. Tran, H. T., Keen, J. D., Krickler, M., Resnick, M. A., and Gordenin, D. A. (1997) *Mol. Cell. Biol.* **17**, 2859–2865
35. Franklin, M. C., Wang, J., and Steitz, T. A. (2001) *Cell* **105**, 657–667
36. Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Ruschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., Hamelin, R., Hamilton, S. R., Hiatt, R. A., Jass, J., Lindblom, A., Lynch, H. T., Peltomaki, P., Ramsey, S. D., Rodriguez-Bigas, M. A., Vasen, H. F., Hawk, E. T., Barrett, J. C., Freedman, A. N., and Srivastava, S. (2004) *J. Natl. Cancer Inst.* **96**, 261–268