Published in final edited form as:

DNA Repair (Amst). 2010 August 5; 9(8): 871–878. doi:10.1016/j.dnarep.2010.05.002.

Mutational clusters generated by non-processive polymerases: a case study using DNA polymerase β *in vitro*

Libertad García-Villada and John W. Drake*

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

Abstract

Available DNA mutational spectra reveal that the number of mutants with multiple mutations (“multiples”) is usually greater than expected from a random distribution of mutations among mutants. These overloads imply the occurrence of non-random clusters of mutations, probably generated during episodes of low-fidelity DNA synthesis. Excess multiples have been reported not only for viruses, bacteria, and eukaryotic cells but also for the DNA polymerases of phages T4 and RB69 *in vitro*. In the simplest case of a purified polymerase, non-random clusters may be generated by a subfraction of phenotypic variants able to introduce more errors per cycle of DNA synthesis than the normal enzyme. According to this hypothesis, excess multiples are not expected with non-processive polymerases even if they harbor rare mutator variants. DNA polymerase β (Pol β) is a mammalian DNA-repair polymerase with very low processivity. Although several Pol β mutational spectra have been described, there is conflicting evidence on whether or not excess multiples occur, with spectra based on the HSV-*tk* system tending to show excess multiples. Excess multiples generated by Pol β or any of its mutants might imply that the excesses of multiples observed in numerous other systems, especially those with processive polymerases, could be artifactual. Here, the distributions of mutations generated by native and recombinant rat Pol β and by the Pol β^{Y265C} mutator were analyzed in the M13mp2 *lacZa* system. Our results present no evidence for a significant excess of multiples over the expected numbers with any of the Pol β enzymes tested in this system. The reported excess of Pol β -generated multiples in the HSV-*tk* system may reflect a reduced efficiency of detection of base substitutions that cause weak phenotypes, which in turn may artifactually increase the frequency of multiples.

Keywords

DNA polymerase β ; mutational clusters; polymerase processivity

1. Introduction

Among mutational spectra, the number of mutants with more than one mutation (“multiples”) is usually greater than expected from the assumption that mutations are randomly distributed among mutants [1]. Overloads of multiples are seen in spectra from

* Corresponding author: John W. Drake, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, 111 South Alexander Drive, Research Triangle Park, NC 27709, USA, Tel: 1 919 541 3361, Fax: 1 919 541 7613, drake@niehs.nih.gov. First author: Libertad García-Villada, garciaivillal@niehs.nih.gov

Conflict of interest statement: The authors declare that there are no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

RNA and DNA viruses, prokaryotes, yeasts, vertebrate cell lines and tissues, and DNA polymerases *in vitro* (reviewed in [2]). These collections of multiples conservatively exclude those that might have occurred as the result of a single error, such as ectopically templated complex mutations and adjacent substitutions where the first mispair may have directly promoted the second mispair. Thus, the observed overloads of multiples imply that mutations are occasionally generated in non-random clusters as the result of some kind of low-fidelity DNA synthesis. Genotypic mutator mutants could in principle be a source of multiples, but this seemed quantitatively unlikely [1,3]. Instead, excess multiples seem to reflect bouts of transient hypermutability, perhaps resulting from errors during the transcription, translation, folding or distribution of proteins involved in DNA replication, and/or by any of numerous other candidate mechanisms ([3], reviewed in [2]). In contrast to random errors in DNA transactions, another source of transient hypermutability is the orchestrated responses of cells to diverse stresses, often eliciting a programmed increase or decrease in specific mutagenic processes [4,5].

Transient hypermutability may enable adaptations that require multiple mutations that are individually neutral or deleterious. Examples would include the adaptation of pathogens to new hosts or to antibiotics. Non-random mutational clusters have also been proposed to promote carcinogenesis [1,6], although the maximum spread of such clusters in mammalian tissues observed to date is approximately the size of a mammalian gene [6] and may thus be insufficient to impact multiple unlinked genes [7].

The simplest system for studying mutational clusters is a DNA polymerase *in vitro*. Excesses of multiples have been observed in mutational spectra from HIV-1 reverse transcriptase, phage T4 DNA polymerase, and the *Escherichia coli* DNA polymerase I Klenow fragment ([1] and references therein), and a striking excess of multiples was described and analyzed in depth for the Pol^{Y567A} Exo⁻ mutator mutant of the phage RB69 DNA polymerase [1]. The existence of these overloads suggests that an ensemble of purified polymerase molecules may contain phenotypic variants with increased error rates [1]. However, to be able to generate an excess of multiples, polymerase variants should not only have an error rate higher than the population average, but also be able to conduct processive DNA synthesis (where processivity is the number of nucleotides inserted during the single cycle between binding and before a polymerase disassociates from the primer-template). Both characteristics must be balanced in such a way that variants have a significantly higher chance of generating more than one error per cycle of DNA synthesis than does a normal polymerase molecule. According to this view, a statistically significant excess of multiples is not expected with non-processive polymerases, even if they harbor rare mutator variants.

DNA polymerase β (Pol β) is a mammalian DNA repair enzyme with very low processivity, sometimes hardly more than one addition per binding cycle [8], although a small proportion of cycles add more nucleotides depending on specific reaction conditions. Due to its small size (39 kDa) and lack of proofreading capacity, Pol β has been used as model in a number of structure-function studies to identify residues that play important roles in the fidelity of its DNA synthesis (reviewed in [9]). As a result, several reports have presented mutational spectra generated *in vitro* by this polymerase or some of its engineered mutator mutants. Two main systems have been used in such studies: phage M13mp2 carrying *lacZa* as a mutation reporter (detailed in [10]), and the HSV-*tk* system (Herpes Simplex Virus type 1 thymidine kinase) [11]. In some descriptions of mutants generated by native and recombinant wild-type Pol β (Pol β^+) using either of these systems, the numbers of multiples were insufficient to support comparisons between observation and expectation [11,12]. In others, mutations that did or did not produce a mutant phenotype were not differentiated [13,14]; as long as only a fraction of all possible mutations (the “detectable” mutations) produce the mutant phenotype when present as single mutations, some way of distinguishing

between detectable and non-detectable mutations is essential for a robust analysis of the distribution of mutations among mutants. On the other hand, mutational data based on the HSV-*tk* system showed such a large number of multiples with a Pol β mutator mutant carrying the Y265C replacement (Pol β^{Y265C}) [12] as to suggest a true excess, a result not expected for a non-processive polymerase. Thus, a paradox arose wherein Pol β did seem to produce non-random mutational clusters.

Because such a paradox would challenge the entire process for defining excess numbers of mutational clusters, we investigated the Pol β matter in depth. We measured processivity and analyzed the distributions of detectable mutations generated by native and recombinant rat Pol β^+ enzymes, as well as the Pol β^{Y265C} mutator mutant, using the M13mp2 *lacZ α* system in which detectable mutations are, in general, well distinguished from mutations that produce no change in phenotype. The possible contribution of a contaminating polymerase in the recombinant Pol β^+ preparations was examined by conducting reactions with a recombinant mutant Pol β (Pol β^{D256A}) that lacks polymerase activity [15].

2. Material and methods

2.1. Enzymes and strains

Pol β^+ from rat Novikoff hepatoma cells was a gift from Dr. Sam Bennett (Oregon State University); this is the same preparation that was used in the primordial Pol β fidelity experiments a quarter of a century ago [13], stored frozen in the meantime. Recombinant rat Pol β^+ and the polymerase-deficient Pol β^{D256A} were gifts from Dr. Sam Wilson's group (NIEHS). Dr. Joann Sweasy's group (Yale University School of Medicine) provided the mutator Pol β^{Y265C} mutant. *Escherichia coli* strains CSH50, MC1061 and NR9099, and wild-type M13mp2 phage have been described in detail elsewhere [10] and were gifts from Dr. Tom Kunkel's group (NIEHS).

2.2. Processivity analysis

Recombinant Pol β^+ and Pol β^{Y265C} processivities were measured using single-stranded wild-type M13mp2 primed with a PAGE-purified 14-mer oligonucleotide (Invitrogen) complementary to positions +(129–115) of the *lacZ α* coding sequence. [γ - 32 P]ATP 5'-primer labeling was carried out with T4 polynucleotide kinase (New England BioLabs) according to the manufacturer's instructions. Labeling reaction mixtures were purified using MicroSpin G-25 Columns (GE Healthcare). Primer hybridization was performed at a 1:1.2 molar ratio of primer:template in SSC by heating to 85°C and slowly cooling to room temperature. Processivity reactions were performed under the same conditions as used in the gap-filling assays (see following section), except that the enzyme:substrate ratio was reduced to 1:40 to ensure that no more than one polymerase bound to most substrate molecules. Samples (5 μ l) from reaction mixtures were removed after 6, 20 and 60 min of incubation, mixed 1:1 with loading buffer (95% deionized formamide, 25 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol) and examined by electrophoresis in a 14% denaturing polyacrylamide gel. Products were quantified by phosphorimager using a Molecular Dynamics Typhoon 9400 and the ImageQuant software.

2.3. In vitro forward mutation assay

The fidelities of native and recombinant Pol β^+ and of Pol β^{Y265C} were analyzed *in vitro* as described [10] using gapped M13mp2 phage as substrate (a 407-nt gap including the coding and upstream sequences of wild-type *lacZ α*). Reaction mixtures (25 μ l) contained 20 mM Tris-HCl (pH 8), 2 mM dithiothreitol (DTT), 25 mM NaCl, 10 mM MgCl₂, 5% glycerol, 1 mM each dNTP (GE Healthcare), and 0.2 nM gapped M13mp2. Pol β was added to a final concentration of 0.1 μ M for Pol β^+ and 0.5 μ M for Pol β^{Y265C} . Reactions were incubated at

37°C for 1 h and stopped with 20 mM EDTA. At least two independent reactions were performed for each enzyme. Complete gap filling was verified by agarose gel electrophoresis. Samples of the reaction products were then electroporated into competent MC1061 cells that were plated with CSH50 cells on minimal agar plates (30 ml) containing IPTG (500 µg) and X-Gal (2.5 mg). Under these conditions, M13mp2 phages without phenotypically detectable mutations in *lacZα* produce dark-blue plaques while phages with phenotypically detectable mutations produce plaques that range from colorless to slightly less than dark blue. Putative mutants were isolated and their phenotypes were verified by plaque purification in CSH50 cells and confirmed mutants were sequenced [10]; rare colorless plaques appear due to mutations located outside of the *lacZα* sequence but are removed from the apparent mutant frequency after sequencing. For Pol β⁺, the sequences of mutants carrying more than one detectable mutation were confirmed in a second round of sequencing. (The detectable mutations have been tabulated [10,16].) Mutation frequencies were converted to rates by dividing by 0.5 instead of the more usual 0.6 because of our own experience with the *lacZα* system.

Error rates, and expected and observed distributions of mutations, were compared using G-tests of independence and of goodness of fit, respectively, and William's correction was always applied [17]. An α value of 0.01 was used in all analyses.

3. Results

3.1. Processivity

The key goal of the present study was to test the hypothesis that a significant excess of mutants carrying multiple mutations is not expected among the DNAs synthesized by a non-processive polymerase such as Pol β. The first step was to check the processivity of Pol β under our experimental conditions. Processivity is a measure of the number of nucleotides a polymerase incorporates into a DNA primer/template per cycle of synthesis (binding, synthesis, and dissociation). To determine processivity *in vitro*, the polymerase:substrate ratio has to be kept low enough to assure that no primed template is copied by more than one polymerase molecule during the course of the reaction. This may be verified by sampling and analyzing the extension products at different times of the reaction: under single-binding conditions, the relative amount of each product should not change with time.

We found such conditions for Pol β⁺ and for Pol β^{Y265C} at a polymerase:substrate ratio of 1:40, where the substrate was ssDNA M13mp2 primed with an oligonucleotide complementary to positions +129 to +115 of the wild-type *lacZα* coding sequence. Figure 1A shows the distribution of extension products generated by these polymerases after 6, 20, and 60 min in three independent reactions. There is a perceptible maximum of ≈ 20 nucleotides added per binding for Pol β⁺ and of ≈ 15 for Pol β^{Y265C}. A representative distribution of products is shown in Figure 1B, where each bar denotes the amount of products of a specific size relative to the sum of all products. Regardless of the differences between Pol β⁺ and Pol β^{Y265C} in the relative amounts of products of each size, both polymerases showed similar overall processivities, usually with 30–40% of the products bearing only one added nucleotide and less than 10% with ≥13 added. Although processivity depends, among other factors, on the specific DNA template sequence, these results signal that both enzymes polymerize according to a common and largely distributive pattern on the *lacZα* template.

3.2. Fidelity

In order to analyze the distribution of mutations generated by native and recombinant Pol β⁺ and the Pol β^{Y265C} mutator, we used the phage M13mp2 *lacZα* forward mutation assay. In

this assay, the M13mp2 double-stranded DNA replicative intermediate with a 407-nt gap comprising the wild-type coding and upstream sequences of *lacZα* is used as a substrate for the synthesis of single-stranded DNA *in vitro*. Products of gap-filling synthesis with detectable mutations in *lacZα* generate a variety of reduced-color plaques in CSH50 cells that contrast with the dark-blue plaques produced by phenotypically wild-type phages. The mutant frequency [(number of observed mutants)/(number of scored plaques)] is estimated after a number of mutants have been confirmed to display a mutant phenotype and have been sequenced.

Mutant frequencies are shown in Table 1. The frequencies generated in two different reactions by native Pol β^+ were 0.089 and 0.118 for an average of 0.104. The frequencies observed for the recombinant Pol β^+ in four different reactions were 0.036, 0.034, 0.037 and 0.025 for an average of 0.033. A 3-fold difference between the frequencies generated by native and recombinant rat Pol β^+ in the M13mp2 *lacZα* forward mutation assay has been observed previously [13,14] and, more generally, differences of this magnitude are expected among Pol β^+ samples from different sources [13]. Gap-filling reactions conducted with Pol β^{Y265C} reflected its mutator activity, producing an average mutant frequency of 0.504 (from two reactions each with frequencies of 0.504). This mutant frequency is close to the upper limit possible in the *lacZα* system and thus cannot be directly compared with the wild-type frequencies, but is at least 5-fold and 16-fold greater than the frequencies observed for the native and recombinant wild-type polymerases, respectively. The frequency of mutants in the control reaction (gapped M13mp2 without polymerase) was 0.0008, which is close to the background frequency for the system when nicked phage DNA is assayed [10,13]; a similar frequency (0.0005) was observed for the reactions conducted with Pol β^{D256A} , and frequencies in this range are usually based on small mutant counts.

A total of 130 mutants generated by native Pol β^+ , 163 mutants generated by recombinant Pol β^+ , and 102 mutants generated by Pol β^{Y265C} were sequenced. Figure 2 shows the mutational spectra for single-base substitutions and indels obtained for native and recombinant Pol β^+ , which are similar to those previously reported for these polymerases [13,14]. Both spectra show the characteristic site specificity of Pol β^+ on the *lacZα* sequence, with two major hot spots: a -T frameshift in a TTTT run at +(70–73) and a T → G transversion at +103. (The slippage mechanism underlying the occurrence of these two hot spots was described previously [18,19].) Sequence changes other than single-base substitutions and indels included 10 larger deletions and 10 complex mutations generated by native Pol β^+ , and 18 larger deletions and 11 complex mutations generated by the recombinant polymerase (Supplementary Tables 1 and 2). Complex mutations include sequence changes that cannot be assigned unambiguously and changes comprising large deletions combined with insertions. Deletions of >1 nt included five and two mutants for native and recombinant Pol β^+ , respectively, with a 317-base deletion (primed by a 5-base direct repeat) that is characteristic of the Pol β^+ mutational spectrum in the *lacZα* system [18,19].

The mutational spectrum obtained from the reactions with Pol β^{Y265C} did not show the same site specificity as the wild-type enzyme, but consisted of both single-base substitutions and indels widely distributed over the *lacZα* sequence (Fig. 3), although Pol β^{Y265C} also preferentially generates single-base deletions in pyrimidine homopolymeric runs. In addition to the single-base changes depicted in Figure 3, the mutational spectrum observed for Pol β^{Y265C} included 41 larger deletions and 9 complex mutations (Supplementary Tables 3 and 4).

Estimated error rates for both Pol β^+ samples were also comparable to those previously reported for these polymerases using the *lacZα* system. As expected, native Pol β^+ showed

overall higher rates than recombinant Pol β^+ (Table 1), reflecting its higher frequency of mutants. However, no significant differences between these two polymerases were found when the observed incidences of the different types of errors (base substitutions, indels of size 1, deletions >1 nt, and complex mutations) or the incidences of all possible base substitutions were compared (Table 2) ($G_{\text{adj}} = 3.66$, d.f. = 3, $P = 0.301$, and $G_{\text{adj}} = 10.83$, d.f. = 5, $P = 0.055$, respectively), indicating that both polymerases have very similar error specificities. All of the error rates estimated for Pol β^{Y265C} were increased over those of the wild-type polymerases (Table 1). Comparing the observed incidences of the different types of errors generated by Pol β^{Y265C} and Pol β^+ (pooled native and recombinant data) showed that these polymerases have different error specificities ($G_{\text{adj}} = 33.70$, d.f. = 3, $P < 0.01$). Adjusted residuals and chi-square partitioning indicated that most of the observed difference was due to a higher frequency of base substitutions generated by the mutator polymerase. Nevertheless, when the incidences of the different types of base substitutions were compared, no significant differences were found ($G_{\text{adj}} = 8.27$, d.f. = 7, $P = 0.309$), suggesting that Pol β^{Y265C} shows an error rate uniformly increased for all types of mispairs. Pol β^{Y265C} also makes more frequent +1 nt indels ($G_{\text{adj}} = 14.44$, d.f. = 1, $P < 0.01$). These insertions were concentrated at a few sites in the mutational spectrum with no apparent preference for homopolymeric runs over non-runs (Fig. 3).

3.3. Distribution of mutations among mutants

Among the 130 *lacZa* mutants isolated from the gap-filling reactions conducted with the native Pol β^+ , 109 contained one detectable mutation, 18 had two, and three had three (Table 3). In the case of the recombinant Pol β^+ , 11 mutants among 163 contained two detectable mutations and no triple was found (Table 3). In addition to these detectable mutations, 29 and 13 non-detectable mutations were observed for the native and recombinant polymerases, respectively, in combination with detectable mutations (Supplementary Table 5). Among 102 sequenced Pol β^{Y265C} mutants, 99 contained two or more detectable *lacZa* mutations, with an average of 5.29 detectable mutations and 10.74 total mutations (detectable + non-detectable) per reporter sequence. Whenever a mutant had more than one mutation per reporter gene, mutations were classified as independent if they were separated by one or more intervening bases or, in the case of tandem mutations (observed only with Pol β^{Y265C}), if any of them appeared alone in any other mutant of the sequenced sample (Supplementary Table 6).

The expected distribution of detectable mutations among mutants is Poisson where the Poisson parameter (μ) was estimated from the sample mean $\Sigma f_a / \Sigma f$, where f_a = the number of reporter sequences carrying a mutations (where $a = 1, 2, \dots$) and $\Sigma f = n$, where n = total reporter sequences analyzed. The number of detectable mutations was 154 for native and 174 for recombinant Pol β^+ . Applying the correction factor 0.5 for the expression of the M13mp2 minus strand (average value under our experimental conditions), the respective μ values were $154/(1268)(0.5) = 0.2429$ and $174/(5079)(0.5) = 0.0685$ detectable mutations per reporter sequence. The expected frequencies for each class of mutant containing k mutations (where $k = 0, 1, 2, \dots$) were estimated as $n(\mu^k e^{-\mu}/k!)$ and then corrected by multiplying by 0.5. Although the number of detectable multiples was slightly greater (< 2-fold) than expected for both polymerases, differences between the expected and the observed distributions of mutations were not significant at $\alpha = 0.01$ ($G_{\text{adj}} = 2.71$, d.f. = 1, $P = 0.100$, and $G_{\text{adj}} = 4.62$, d.f. = 1, $P = 0.032$ for native and recombinant Pol β^+ , respectively). The possibility that some of the observed multiples were produced by a contaminating polymerase seemed unlikely because no multiples were found with the polymerase-deficient Pol β^{D256A} , whose mutational spectrum was indistinguishable from the background (data not shown). On the other hand, if the observed excess of multiples was due to a subpopulation of mutator polymerases, one might anticipate a skewed distribution

of distances between consecutive mutations, with a significantly higher-than-expected number of observations in the interval of processivity values showed by the enzyme population. However, when observed and expected distributions of distances between consecutive mutations (both detectable and non-detectable) were compared (Fig. 4), no significant differences were found (native and recombinant data pooled, $G_{\text{adj}} = 2.93$, d.f. = 6, $P = 0.818$). These results, together with an observed excess of non-mutants for both polymerases (Class 0 in Table 3), suggest that discrepancies between the expected and the observed distributions of mutations are more probably due to a failure to detect a few singles with weak phenotypes than to a real excess of multiples (see Discussion).

In the case of Pol β^{Y265C} , when the corresponding expected distribution of detectable mutations was compared with the observed distribution (Fig. 5), no significant differences were found ($G_{\text{adj}} = 6.30$, d.f. = 8, $P = 0.614$). Furthermore, the observed and expected numbers of multiples were the same (99 versus 98.76, respectively).

4. Discussion

Errors during the transcription and translation of proteins are estimated to generate $(1-5) \times 10^{-4}$ errors per amino acid (reviewed in [20]). For Pol β 's 335 amino acids, these rates imply that about 1 in 20 molecules will contain a non-heritable amino acid substitution. In addition, proteins can misfold in diverse ways, sometimes producing an active but perturbed enzyme [21]. If any of these alterations affect the active site of a polymerase, fidelity may be compromised. Thus, any population of polymerase molecules probably contains a subpopulation of diverse variants with increased error rates. When synthesizing DNA processively, this subpopulation may generate an excess of products carrying more than one mutation. Examples of this phenomenon have previously been tabulated [1]. However, this overload of multiples is only possible if processivity is high enough to allow such phenotypic mutators to make more multiple mistakes per DNA binding than the normal polymerase. According to this logic, non-processive polymerases such as Pol β^+ and Pol β^{Y265C} are expected not to generate significant excesses of multiples even if their populations contain phenotypic mutators. The opposite might imply that the excesses of multiples observed with processive polymerases such as HIV-1 RT and RB69 Pol^{Y567A} Exo⁻ are artifactual. To test this hypothesis, we first checked the processivity of both Pol β^+ and Pol β^{Y265C} using the same DNA *lacZa* substrate that we used to analyze the mutational spectra of these enzymes. We found that both polymerases have a probability of dissociating from a DNA substrate of roughly 30–40% after introducing one nt, rising to about 80% after having introduced five (summing the percentages of products ≤ 5 nt long in Fig. 1). These results agree with those previously reported for these polymerases in the HSV-*tk* system [12] as well as with those reported for Pol β^+ in other recessed DNA templates [8,22–28] and indicate that Pol β is a strongly distributive enzyme that introduces no more than a few nucleotides per binding to a primed ssDNA template. This result implies that its chance of generating more than one independent error per binding is probably negligible.

The fidelity results that we obtained with native and recombinant Pol β^+ were similar overall to those previously reported in the M13mp2 *lacZa* system [13,14]. There is no previous report on the fidelity of Pol β^{Y265C} using *lacZa*. Results from the HSV-*tk* system indicate that Pol β^{Y265C} is a strong mutator for both base substitutions and indels [12]. The mutator factor we observed for Pol β^{Y265C} in the *lacZa* system is similar to that reported in the HSV-*tk* system, with an average mutation frequency about 76 times that of the recombinant wild-type enzyme and high rates for both base substitutions and indels. The role of tyrosine 265 in Pol β fidelity has been analyzed previously [12,29]. In synthesis, this residue seems to mediate a conformational change that provides selection against premutational intermediates, so that the Y265C substitution reduces the overall fidelity.

When we analyzed the distributions of detectable mutations generated by Pol β^+ and Pol β^{Y265C} , we found no significant differences between the observed and expected distributions, that is, the mutations appeared to be Poisson-distributed, as would be expected if they were generated independently. These results are in agreement with the processivity values that we observed for both polymerases and indicate that mutations in multiples were mostly generated in separate DNA-binding events. Although we saw a small excess of multiples with Pol β^+ in the *lacZa* system, our analysis suggested that this excess, if real, could have been caused by occasional failures to detect mutations with weak phenotypes when present as singles, which would have led to an underestimate of the mutant frequency (and thus the mutation rate), but no failures to detect them in combination with other detectable mutations. In agreement with this explanation, we found no differences between the observed and expected numbers of multiples with Pol β^{Y265C} , which generates mostly multiples (97% of total mutants) in the *lacZa* system, thus giving little chance for failing to detect singles. Among the factors that may inflate the frequency of multiples, underestimating the overall (singles + multiples) mutant frequency may be one of the more probable when using systems in which some base substitutions generate a weak mutant phenotype that is difficult to detect [2]. In the *lacZa* system, for instance, base substitutions generating mutant plaques only slightly less dark than wild-type plaques can cause the mutant frequency to be underestimated. Nevertheless, our results indicate that such underestimation, if it occurred, was insufficient to generate significant differences between the observed and the expected distributions of mutations.

An apparent excess of multiples produced by Pol β^{Y265C} in the HSV-*tk* system was reported in Opresko et al. [12]. In a sample of 79 sequenced mutants, while the expected number of doubles (containing two detectable mutations) was 2.1 (estimated from the mutant frequency and the number of mutations generated by Pol β^{Y265C} in [12]), 10 were recorded. This 5-fold difference contrasts with our results because we found no significant differences between the observed and the expected distributions of detectable mutations generated by this polymerase in the *lacZa* system. If a preparation of Pol β^{Y265C} contained mutator variants, non-random clusters of mutations would be constrained by the processivity of this polymerase, and thus excess multiples would be expected to contain mutations mostly separated by no more than 15 nt. However, among the 10 doubles observed for Pol β^{Y265C} in the HSV-*tk* system, nine contained mutations separated by more than 47 nt, suggesting that they were generated in independent polymerase binding events as, in fact, was suggested by the authors [12]. Thus, if mutations in multiples were generated independently and there was a clear excess of multiples over the expected according to a random distribution of mutations among mutants, some factor other than phenotypic mutator polymerases must have generated the Pol β^{Y265C} -generated doubles in the HSV-*tk* reporter.

Although mutational data generated using different reporters often cannot be directly compared, a clue about the differences between the present study and that of Opresko et al. may be provided by the coefficients of variation (CV) of mutant frequencies observed in each system, where $CV = 100(\text{standard deviation of mutant frequency})/(\text{average mutant frequency})$. In the *lacZa* system, the CV for recombinant Pol β^+ is $(100)(0.006)/0.033 = 18$. In the case of Pol β^{Y265C} , which generates mutants (mostly multiples) at a frequency at the upper limit of detection with the *lacZa* system detection, thus allowing little variation between assays, we would anticipate a $CV \leq 10$ because standard deviations may be up to 20% of averaged mutant frequencies when multiple determinations are performed using the M13mp2 *lacZa* system [30]. In the HSV-*tk* system, Pol β^+ has a $CV = (100)(0.0008)/0.0032 = 25$, and Pol β^{Y265C} has a $CV = (100)(0.02)/0.073 = 27$ (data from [12]). These high CV values observed for the mutant frequencies reported in the HSV-*tk* system suggest greater fluctuations in the efficiency of mutant detection in this system than in the *lacZa* system. These fluctuations may reflect difficulties in detecting base substitutions that generate weak

mutants, as may also sometimes happen in the *lacZα* system. If mutations with a very weak phenotype as singles gained a strong phenotype as doubles, an artifactual excess of multiples could result. This phenomenon may occur in the HSV-*tk* system, where leaky mutants can be killed by the selection process [11]. Thus, while the reported frequency for indels generated by Pol β^{Y265C} in the HSV-*tk* system is 4.5-fold greater than the frequency of base substitutions, four of the 10 doubles reported for this polymerase were composed of a detectable base substitution and a detectable indel, again suggesting that a significant fraction of base-pair substitutions was not detected as single mutations.

Altogether, our results present no evidence for a significant excess of multiples with Pol β⁺ or Pol β^{Y265C} in the *lacZα* system. These results support the hypothesis that an excess of multiples can only be generated processively in a hypermutated tract, and thus support the reality of the excesses of multiples observed in the *lacZα* system for other (processive) polymerases [1]. The reported excess of multiples with Pol β^{Y265C} in the HSV-*tk* system seems to be simple consequence of a somewhat low efficiency of detection of base substitutions with weak phenotypes, a trait that does not, however, affect numerous other conclusions reached with that system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Sam Bennett, the Sam Wilson lab, and Joann Sweasy for gifts of purified Pol β preparations; Kasia Bebenek, Stephanie Nick McElhinny, Dinh Nguyen and Tom Kunkel for advice on the properties and use of the *lacZα* system; and Kasia Bebenek and Bill Beard for critical readings of the manuscript. This research was supported by funds allocated to project number Z01ES065016 of the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

References

1. Drake JW, Bebenek A, Kissling GE, Peddada S. Clusters of mutations from transient hypermutability. *Proc Natl Acad Sci USA*. 2005; 102:12849–12854. [PubMed: 16118275]
2. Drake JW. Too many mutants with multiple mutations. *Crit Rev Biochem Mol Bio*. 2007; 42:247–258. [PubMed: 17687667]
3. Ninio J. Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics*. 1991; 129:957–962. [PubMed: 1752431]
4. Galhardo RS, Hastings PJ, Rosenberg SM. Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Bio*. 2007; 42:399–435. [PubMed: 17917874]
5. Gonzalez C, Hadany L, Ponder RG, Price M, Hastings PJ, Rosenberg SM. Mutability and importance of a hypermutable cell subpopulation that produces stress-induced mutants in *Escherichia coli*. *PLoS Genet*. 2008; 4:e1000208. [PubMed: 18833303]
6. Wang J, Gonzalez KD, Scaringe WA, Tsai K, Liu N, Gu D, Li W, Hill KA, Sommer SS. Evidence for mutation showers. *Proc Natl Acad Sci USA*. 2007; 104:8403–8408. [PubMed: 17485671]
7. Drake JW. Mutations in clusters and showers. *Proc Natl Acad Sci USA*. 2007; 104:8203–8204. [PubMed: 17495029]
8. Bambara RA, Uyemura D, Choi T. On the processive mechanism of *Escherichia coli* DNA polymerase I. Quantitative assessment of processivity. *J Biol Chem*. 1978; 253:413–423. [PubMed: 338607]
9. Beard WA, Wilson SH. Structure and mechanism of DNA polymerase β. *Chem Rev*. 2006; 106:361–382. [PubMed: 16464010]
10. Bebenek K, Kunkel TA. Analyzing fidelity of DNA polymerases. *Meth Enzymol*. 1995; 262:217–232. [PubMed: 8594349]

11. Eckert KA, Hile SE, Vargo PL. Development and use of an *in vitro* HSV-*tk* forward mutation assay to study eukaryotic DNA polymerase processing of DNA alkyl lesions. *Nucl Acids Res.* 1997; 25:1450–1457. [PubMed: 9060443]
12. Opreko PL, Sweasy JB, Eckert KA. The mutator form of polymerase β with amino acid substitution at tyrosine 265 in the hinge region displays an increase in both base substitution and frame shift errors. *Biochem.* 1998; 37:2111–2119. [PubMed: 9485358]
13. Kunkel TA. The mutational specificity of DNA polymerase- β during *in vitro* DNA synthesis. Production of frameshift, base substitution, and deletion mutations. *J Biol Chem.* 1985; 260:5787–5796. [PubMed: 3988773]
14. Osheroff WP, Jung HK, Beard WA, Wilson SH, Kunkel TA. The fidelity of DNA polymerase β during distributive and processive synthesis. *J Biol Chem.* 1999; 274:3642–3650. [PubMed: 9920913]
15. Menge KL, Hostomsky Z, Nodes BR, Hudson GO, Rahmati S, Moonmaw EW, Almasy RJ, Hostomska Z. Structure-function analysis of the mammalian DNA polymerase β active site: role of aspartic acid 256, arginine 254, and arginine 258 in nucleotidyl transfer. *Biochem.* 1995; 34:15934–15942. [PubMed: 8519750]
16. Bebenek A, Carver GT, Dressman HK, Kadyrov FA, Haseman JK, Petrov V, Konigsberg WH, Karam JD, Drake JW. Dissecting the fidelity of bacteriophage RB69 DNA polymerase: site-specific modulation of fidelity by polymerase accessory proteins. *Genetics.* 2002; 162:1003–1018. [PubMed: 12454051]
17. Sokal, RR.; Rohlf, FJ. *Biometry.* third. Freeman & Co.; New York: 1995.
18. Kunkel TA, Alexander PS. The base substitution fidelity of eukaryotic DNA polymerases. *J Biol Chem.* 1986; 261:160–166. [PubMed: 3941068]
19. Kunkel TA, Soni A. Mutagenesis by transient misalignment. *J Biol Chem.* 1988; 263:14784–14789. [PubMed: 3049589]
20. Willensdorfer M, Bürger R, Nowak MA. Phenotypic mutation rates and the abundance of abnormal proteins in yeast. *PLoS Comput Biol.* 2007; 3:e203. [PubMed: 18039025]
21. Stöhr J, Weinmann N, Wille H, Kaimann T, Nagel-Steger L, Birkmann E, Panza G, Prusiner SB, Eigen M, Reisner D. Mechanisms of prion protein assembly into amyloid. *Proc Natl Acad Sci USA.* 2008; 105:2409–2414. [PubMed: 18268326]
22. Chang LMS. The distributive nature of enzymatic DNA synthesis. *J Mol Biol.* 1975; 93:219–235. [PubMed: 239246]
23. Das SK, Fujimura RK. Processiveness of DNA polymerases. A comparative study using a simple procedure. *J Biol Chem.* 1979; 254:1227–1232. [PubMed: 368069]
24. Detera SD, Becerra SP, Swack JA, Wilson SH. Studies on the mechanism of DNA polymerase α . Nascent chain elongation, steady state kinetics, and the initiation phase of DNA synthesis. *J Biol Chem.* 1981; 256:6933–6943. [PubMed: 7240254]
25. Abbotts J, Sen Gupta DN, Zmudzka B, Widen SG, Notario V, Wilson SH. Expression of human DNA polymerase β in *Escherichia coli* and characterization of the recombinant enzyme. *Biochem.* 1988; 27:901–909. [PubMed: 3284575]
26. Werneburg BG, Ahn J, Zhong X, Hondal RJ, Kraynov VS, Tsai MD. DNA polymerase β : pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity. *Biochem.* 1996; 35:7041–7050. [PubMed: 8679529]
27. Chagovetz AM, Sweasy JB, Preston BD. Increased activity and fidelity of DNA polymerase β on single-nucleotide gapped DNA. *J Biol Chem.* 1997; 272:27501–27504. [PubMed: 9346877]
28. Singhal RK, Wilson SH. Short gap-filling synthesis by DNA polymerase beta is processive. *J Biol Chem.* 1993; 268:15906–15911. [PubMed: 8340415]
29. Washington SL, Yoon MS, Chagovetz AM, Li SX, Clairmont CA, Preston BD, Eckert KA, Sweasy JB. A genetic system to identify DNA polymerase β mutator mutants. *Proc Nat Acad Sci USA.* 1997; 94:1321–1326. [PubMed: 9037051]
30. Longley MJ, Nguyen D, Kunkel TA, Copeland WC. The fidelity of human DNA polymerase γ with and without exonucleolytic proofreading and the p55 accessory subunit. *J Biol Chem.* 2001; 276:38555–38562. [PubMed: 11504725]

Figure 1. Distribution of products during a single cycle of DNA synthesis by recombinant Pol β⁺ and Pol β^{Y265C} at *lacZα*. (A) Phosphor image of the reaction products of processive synthesis separated on a 14% denaturing polyacrylamide gel. Lanes B, C and D represent primer extension products generated by Pol β⁺ at 6, 20 and 60 min in three different reactions. Lanes F, G and H represent primer extension products generated by Pol β^{Y265C} at 6, 20 and 60 min in three different reactions. Reactions were performed as described in the Material and Methods. Lanes A and E represent control reactions without enzyme. (B) Representative diagram showing the distribution of extended products. Plotted values are the average and SD of three different reactions.

Figure 2.

Spectrum of mutations generated by Pol β^+ during gap-filling synthesis *in vitro*. The DNA sequence in the figure represents the template strand of the *lacZa* reporter (wild-type coding and upstream sequences of the *lacZa* gene) carried by M13mp2. Bases on blue backgrounds mark decamers and the base on a red background indicates the first transcribed nucleotide of the *lacZa* complementation region. Mutations produced by native and recombinant Pol β^+ are indicated above and below the sequence, respectively. Triangles represent single-base deletions, while inverted triangles represent single-base additions. Triangles are depicted closed for indels arising in homopolymeric runs. Detectable mutations are shown in red and undetectable mutations in black. Complex mutations and deletions comprising more than one nt are shown in Supplementary Tables 1 and 2.

Figure 3.

Spectrum of mutations generated by Pol β^{Y265C} during gap-filling synthesis *in vitro*. Base substitutions and indels are shown above and below the DNA substrate sequence, respectively. See legend to Figure 2 for other details. Complex mutations and deletions comprising more than one nt are shown in Supplementary Tables 3 and 4.

Figure 4.

Observed and expected distributions of distances (in bp) between consecutive mutations (detectable and not detectable) generated by Pol β^+ (pooling native and recombinant enzyme data) during gap-filling synthesis *in vitro*. The expected distribution of distances was determined from the mutational spectra generated by Pol β^+ , calculating the distances between all observed mutations (taken 2-by-2, so that 4 mutations provide 3 distances) and correcting the resultant frequencies for the total number (76) of observations.

Figure 5. Observed and expected distributions of detectable mutations among *lacZα* sequences synthesized by Pol β^{Y265C} *in vitro*. See text for details about estimating the expected frequencies.

Table 1

Fidelities of native and recombinant Pol β^+ and of Pol β^{Y265C} in the *lacZa* system

	Native Pol β^+	Recombinant Pol β^+	Pol β^{Y265C}
Total plaques	1,268	5,079	6,773
Mutant frequency	0.104	0.033	0.504
<i>lacZa</i> mutants sequenced	130	163	102
Detectable <i>lacZa</i> mutations	154	174	540
Mutation frequency ^a	0.243	0.069	5.31
Error rates (per 10 ⁴) ^b			
Base substitutions	8.36 (69)	1.87 (62)	235 (312)
Indels (−1)	4.92 (62)	1.547 (78)	71.0 (144)
Indels (+1)	0.24 (3)	0.10 (5)	16.8 (34)
Frequency ^c of deletions >1 nt	0.022 (14)	0.011 (27)	0.401 (41)
Frequency ^c of complex mutations ^d	0.009 (6)	0.0008 (2)	0.088 (9)

^a (*lacZa* mutants)/(mutations in sequenced mutants)/0.5(total plaques)(sequenced mutants).

^b Specific error rate per reporter sequence, calculated for phenotypically detectable mutations as described [10] using the divisor 0.5 to convert from frequencies and normalizing to numbers of sites at which that class of mutations could arise. Numbers in parentheses are the number of times each kind of mutation was observed.

^c Frequencies were calculated as (number of times the kind of mutation was observed)/0.5(total plaques). For Pol β^{Y265C} , the total number of plaques to be considered in these calculations was estimated as (*lacZa* mutants sequenced)/(mutant frequency).

^d Complex mutations include sequence changes that cannot be assigned unambiguously and large deletions combined with insertions (see text).

Table 2
Base substitution error rates for native and recombinant Pol β^+ and for Pol β^{Y265C}

Base substitution	Mispair	Native Pol β^+	Recombinant Pol β^+	Pol β^{Y265C}
A→G	A-dCMP	5.8 (7)	0.2 (1)	170(33)
A→T	A-dAMP	1.3 (2)	0.5 (3)	63(16)
A→C	A-dGMP	≤0.9 (0)	≤0.2 (0)	5.8 (1)
G→A	G-dTMP	9.4 (13)	2.2 (12)	290(66)
G→C	G-dGMP	≤0.8 (0)	0.2 (1)	15(3)
G→T	G-dAMP	1.9 (3)	1.3 (8)	75(19)
C→T	C-dAMP	3.8 (6)	2.2 (14)	140(36)
C→G	C-dCMP	≤1.8 (0)	≤0.4 (0)	22(2)
C→A	C-dTMP	0.9 (1)	≤0.2 (0)	40(7)
T→C	T-dGMP	7.9 (14)	1.3 (9)	240(68)
T→A	T-dTMP	3.0 (3)	≤0.3 (0)	25(4)
T→G	T-dCMP	13(20)	2.3 (14)	230(57)

Error rates $\times 10^4$ for individual base substitutions were calculated as described in Table 1. Numbers in parentheses represent number of times each kind of base substitution was observed; rates for 0 mutations are $\leq$ the rate for 1 mutation.

Expected and observed distributions of detectable mutations among *lacZα* sequences synthesized by native and recombinant rat Pol β⁺ *in vitro*

Table 3

No. of mutations	Native Pol β ⁺		Recombinant Pol β ⁺	
	Expected	Observed	Expected	Observed
0	1132.7	1138	4907.3	4916
1	119.7	109	165.7	152
2	14.4	18	5.8	11
3	1.1	3	0.14	0
≥4	0.074	0	0.002	0
Total	1268	1268	5079	5079

See text concerning the estimation of the expected frequencies.